## REMARKS

### I. Status of the Application

Claims 1-3, 6, 12, 14, 15, 52-54, 56-60, 70, 71, and 73-76 are presently pending in the application. Applicants thank the Examiner for withdrawing the previous objections denoted "Q, R, and S" in the previous Office Action. Applicants further thank the Examiner for withdrawing the previously asserted new matter and written description rejections.

Claims 1-3, 6, 12, 14-15, 52-60, and 70-76 stand rejected under 35 U.S.C. § 112, ¶2. Claims 1-2, 6, 12, 15, 52, 53, 56, 57, 70-71, and 73 stand rejected under 35 U.S.C. § 102(b) over US 5,242,974 ("Holmes"). Claims 1-3, 12, 52-54, 57, 60, 70-71, and 76 stand rejected under 35 U.S.C. § 102(b) over MacDonald et al., *Acc. Chem. Res.*, 27(6):151 (1994) ("MacDonald").

Further, the Examiner is presently applying the following new rejections. Claims 1-3, 6, 12, 14-15, 52-54, 56-60, 70-71, and 73-76 stand rejected under 35 U.S.C. § 112, ¶1 pursuant to the written description requirement. Claims 1-3, 6, 12, 14-15, 52-54, 56-60, 70-71, and 73-76 stand rejected under 35 U.S.C. § 112, ¶1 pursuant to the enablement requirement. Claims 1-3, 6, 12, 14-15, 52-54, 56-60, 70-71 and 73-76 stand rejected under 35 U.S.C. § 112, ¶2. Claims 1-2, 6, 12, 15, 52-54, 56-60, 70-71, and 73-76 stand rejected under 35 U.S.C. § 102(b) over US 5,143,854 ("Pirrung"). Claims 1-3, 6, 12, 14-15, 52-54, 56-60, 70-71, and 73-76 stand rejected under the obviousness-type double patenting doctrine as to US Publication No. 2002/0022721 ("the '721 publication"). Claims 1-3, 6, 12, 14-15, 52-54, 56-60, 70-71, and 73-76 stand rejected under the obviousness-type double patenting doctrine as to US 6,083,697 ("the '697 patent"). Applicants respectfully traverse all of the foregoing rejections.

## II. Claims 1-3, 6, 12, 14-15, 52-60, and 70-76 Are Definite

Claims 1-3, 6, 12, 14-15, 52-60, and 70-76 stand rejected under 35 U.S.C. § 112, ¶2 as being indefinite as to the term "masked acids." Applicants respectfully traverse this rejection in view of the following remarks.

The Examiner's present rejection at pages 3-4 of the present Office Action is limited to the term "masked acids", which appears only in dependent claims 3 and 54. As such, Applicants respectfully request that the present rejection be limited to claims 3 and 54 and not claims 1-3, 6, 12, 14-15, 52-60, and 70-76, as indicated by the Examiner.

Nevertheless, claims 3 and 54 are definite insofar as the term "masked acids" is definite. The Examiner asserts that the term "masked acids" is defined only by their function, i.e., their ability to function as an autocatalyst. Applicants respectfully disagree because the term "masked acids" is readily recognizable by those of skill in the art because this term is used by those of skill in the art. For example, the term "masked acid" appears in issued patents, such as parent to the present application, U.S. Pat. No. 6,083,697 (column 14, lines 61-63) and U.S. Pat. No. 6,482,954 (column 7, lines 44-46), which are seen at Tabs A and B, respectively; pending patent applications, such as U.S. Pub. No. 2003/0092033 (paragraph 17), which is seen at Tab C; journals, such as Organic Letters (Owen et al., *Organic Letters*, 4(14): 2293, 2294 (2002), which is seen at Tab D. Further, the term "masked acid" is used by professors of organic synthesis (see Tab E) and major vendors of chemical compounds (see Tab F).

Applicants note in parent U.S. Pat. No. 6,083,697, which issued July 4, 2000, the term "masked acid" is defined by claim 3.

Further, Applicants describe masked acids throughout the specification, for example, at page 20, lines 28-29: "In some preferred embodiments of polynucleotide synthesis, masked acids

including esters, anhydrides, and nitrites are used as autocatalysts." Based on the evidence presented above, it is apparent that those of skill in the art understand that esters, anhydrides, and nitrites can serve as masked acids because of the structural features of these classes of compounds. For example, those of skill in the art understand that esters (-COOR) can be masked acids since the R group protects (i.e., masks) the acid (-COOH). Thus, esters, because of their structural resemblance to acids are suitable masked acids. Applicants note that masked acids are thus definite because of their *structural* features, and not by their function, as asserted by the Examiner. Thus, removal of the present rejection is respectfully requested.

## III. Claims 1-2, 6, 12, 15, 52, 53, 56, 57, 70-71, 73 Are Patentable over Holmes

Claims 1-2, 6, 12, 15, 52, 53, 56, 57, 70-71, and 73 stand rejected under 35 U.S.C. § 102(b) over Holmes. Applicants respectfully traverse this rejection.

Independent claims 1, 52, and 70 are patentable over Holmes at least because Holmes fails to disclose a radiation sensitive compound or a photosensitive acid compound, each producing a catalyst when irradiated. Similarly, Holmes fails to disclose an autocatalytic compound or group generating a protecting group removing product, as defined by all pending independent claims. Instead, Holmes teaches a method for cyclization and reversal of the polarity of polymers on a substrate.

The Examiner asserts at page 5 of the Office Action, "the NVOC protection group can function as a radiation sensitive compound because NVOC can function as a catalyst (6,6-bisveratric acid catalyzes the removal of the NVOC protecting group) and an autocatalyst (6,6-bisveratric acid catalyzes the formation of more 6,6-bisveratric acid) because the NVOC protecting group is an acid-labile protecting group..." The Examiner then cites column 19, lines

29-30 to show that Holmes allegedly shows the susceptibility of NVOC to acid cleavage. Applicants respectfully disagree with this assertion.

At the outset, Applicants note that Holmes at column 19, lines 29-30 merely shows the general formula defined by claim 17. Nowhere does Holmes disclose a photosensitive compound that produces a catalyst when irradiated. Further, nowhere does Holmes disclose an autocatalytic compound or group generating a protecting group removing product. The Examiner appears to be making this assertion based on the teachings of previously cited US 5,679,773, which is not currently being cited.

The Examiner asserts that Applicants bear the burden of showing the differences between the presently claimed compounds and the compounds cited in the references. Applicants believe that they satisfied this burden, notably at pages 10-13 in the previous Response dated June 20, 2003, which is incorporated here by reference, by explaining certain distinguishing characteristics of its photosensitive compounds and autocatalytic compounds vis-à-vis Holmes' NVOC group. However, for the Examiner's convenience, the following summarizes certain distinguishing characteristics:

1. Applicants' photosensitive compounds produce a catalyst when irradiated. Holmes does not disclose that his NVOC protecting group catalyzes anything after dissociating from the compound it is protecting.
2. Applicants' disclose at page 11, lines 25-26 that an autocatalyst "is a material that undergoes a reaction that produces a product that is also a catalyst for that same reaction." Holmes does not disclose that his NVOC protecting group is a catalyst for any reaction, let alone a reaction that produces more NVOC.

The Examiner asserts at page 5 of the Office Action that NVOC is cleaved into 6,6-bisveratric acid by light, which supposedly catalyzes the removal of the NVOC group as well as the formation of more 6,6,-bisveratric acid. However, the Examiner has not indicated anywhere in Holmes (or provided any other evidence) where such a concept is disclosed. Indeed, Holmes does not explicitly disclose that NVOC produces a catalyst upon being irradiated, as required by claims 1, 52, and 70. Further, Holmes does not explicitly disclose that NVOC is a material that undergoes a reaction that produces a product that is also a catalyst for that same reaction, i.e., is an autocatalyst, as required by all pending independent claims. It is also worth noting that NVOC is an amine protecting group – not an acid protecting group. See Spivey, A.C. and Andrews, B.I., *Annu. Rep. Prog. Chem.*, 96(Sect. B.): 65, 77-78 (2000) at Tab G. Thus, removal of the NVOC group from the compound it protects would produce a free amine – not an acid. If the Examiner believes that Holmes inherently discloses that NVOC is a catalyst, then the Examiner is respectfully reminded that he bears the burden of proving rationale or evidence showing such concepts. MPEP § 2122.

In view of the foregoing, Applicants have provided independently sufficient bases for the patentability of claims 1, 52, and 70 and all claims dependent therefrom over Holmes. However, to be fully responsive to the present Office Action, the following comments are provided.

Claim 2 is patentable over Holmes because Holmes does not disclose a photosensitive compound or group *that produces a catalyst when irradiated.* As discussed immediately above, nowhere does Holmes disclose that the NVOC group catalyzes any reaction.

Claim 12 is patentable over Holmes because Holmes does not disclose a catalyst, let alone a photoactivated catalyst or a photoactivated acid catalyst, as defined by claim 12. Applicants state at page 14, lines 15-18: "In certain embodiments, a photo activated acid catalyst

(PAAC) is irradiated. The resulting acid produced from the PAAC activates an enhancer to undergo an acid-catalyzed reaction to itself produce an acid that removes acid labile protecting groups from a linker molecule or synthesis intermediate." Certainly, Holmes does not disclose (and the Examiner has not shown with support) that NVOC activates an enhancer to undergo an acid-catalyzed reaction to itself produce an acid that removes acid labile protecting groups. Further, Holmes does not disclose toluenesulfonic acid, as defined by claim 12.

Claims 53 and 71 are patentable over Holmes because Holmes does not disclose a photoactivated acid catalyst, as discussed immediately above.

Claims 56, 57, and 73 are patentable over Holmes at least by virtue of its dependency from claim 52, which Applicants demonstrated above is patentable. Accordingly, removal of the present rejection is respectfully requested at this time.

## IV. Claims 1-3, 12, 52-54, 57, 60, 70-71, and 76 Are Patentable over MacDonald

Claims 1-3, 12, 52-54, 57, 60, 70-71, and 76 stand rejected under 35 U.S.C. § 102(b) over MacDonald. Applicants respectfully traverse this rejection.

Independent claim 1 defines a catalyst system comprising: (1) a radiation sensitive compound that upon being irradiated produces; (2) a catalyst; and (3) an autocatalytic compound that upon being activated by the catalyst generates; (4) a protecting group removing product. Independent claim 52 defines a catalyst system comprising: (1) a photosensitive acid compound that upon being irradiated produces; (2) a catalyst; and (3) an autocatalytic compound that upon being activated by the catalyst generates; (4) a protecting group removing product. Similarly, Independent claim 70 as amended defines (1) a synthesis intermediate having; (2) an acid removable protecting group; (3) a photosensitive acid, compound or group that upon irradiation

produces; (4) an acid; and (5) an autocatalytic compound that upon activation by the acid generates; (6) a protecting group removing product.

Independent claims 1, 52, and 70 and all claims dependent therefrom are patentable over MacDonald at least because MacDonald fails to disclose an autocatalytic compound, which is required by each of the independent claims. The Examiner asserts at page 7 of the present Office Action that MacDonald discloses: (1) a radiation sensitive compound (i.e., triphenylsulfonium hexafluoroantimonate); (2) a catalyst (i.e., $H^+$, referring to $H^+CF_3SO_3^-$); (3) an autocatalytic compound (i.e., t-BOC group); and (4) a protecting group removing product (i.e., t-butyl cation).

As Applicants explained above, an autocatalytic compound is defined in the specification at page 11, lines 25-26 as "a material that undergoes a reaction that produces a product that is also a catalyst for that same reaction." The t-BOC group in MacDonald does not undergo a reaction that produces a product that is also a catalyst for that same reaction. For t-BOC to be an autocatalytic compound, the t-BOC group would have to undergo a reaction that produces a product (i.e., t-butyl cation, as asserted by the Examiner) that catalyzes the reaction for removing the t-BOC group. However, this is simply not the case. MacDonald makes it clear at page 152, column 2 that $H^+CF_3SO_3^-$ instead catalyzes the removal of the t-BOC group:

> Therefore the "t-BOC resist" system consists of a mixture of poly[4-[(tert-butyloxycarbonyl)oxy styrene] (PBOCST), a polymer with labile tert-butyl carbonate side groups, and **triphenylsulfonium hexafluoroantimonate, a photoactive composition that generates strong acid**. Figures 2 and 3 outline the mechanism of operation of this resist. Exposing the resist film through a mask (Figure 2) creates a latent image consisting of exposed areas with strong acid dispersed in the tert-butyl carbonate-containing polymer along with unchanged unexposed areas. Upon heating to 100°C, **the acid catalyzes the decomposition of the t-BOC groups** with release of $CO_2$ and isobutylene (Figure 3), reducing the resist film thickness in the exposed areas, and modifying the chemical composition of the coating. (Emphasis added).

Certainly, when MacDonald states that the acid catalyzes the decomposition of the t-BOC groups, he is referring to the strong acid generated by triphenylsulfonium hexafluoroantimonate, i.e., $H^+CF_3SO_3^-$. Thus, MacDonald's disclosure of the t-butyl cation is not a disclosure of an autocatalytic compound.

Further, it is worth mentioning that $H^+CF_3SO_3^-$ is not a disclosure of an autocatalyst because this acid is not catalyzing the reaction to produce more $H^+CF_3SO_3^-$. At best, $H^+CF_3SO_3^-$ catalyzes a different reaction, i.e., removal of the t-BOC groups. Similarly, triphenylsulfonium hexafluoroantimonate is not an autocatalytic compound. The Examiner concedes this point at page 9 of the present Office Action. Thus, claim 1 is patentable over MacDonald at least because MacDonald does not disclose an autocatalytic compound.

In view of the foregoing, Applicants have provided independently sufficient bases for the patentability of claims 1, 52, and 70 and all claims dependent therefrom over MacDonald. However, to be fully responsive to the present Office Action, the following comments are provided.

Claims 2, 12, 53, 57, 60, 71, and 76 are patentable over MacDonald at least by virtue of their direct or indirect dependency from claims 1, 52, and 70, which Applicants demonstrated above are patentable.

Claims 3 and 54 are patentable over MacDonald at least because MacDonald does not disclose an autocatalytic compound. Further, MacDonald's t-BOC is not a disclosure of a masked acid. t-BOC is a protecting group. See page 10, lines 9-31 (quoted below).

## V. Claims 1-3, 6, 12, 14-15, 52-54, 56-60, 70-71, and 73-76 Satisfy the Written Description Requirement

Claims 1-3, 6, 12, 14-15, 52-54, 56-60, 70-71, and 73-76 stand rejected under 35 U.S.C. § 112, ¶1 pursuant to the written description requirement. Applicants traverse this rejection in view of the following remarks.

The Examiner asserts that the following terms defined by the pending claims have no distinguishing structural attributes: protecting group, linker molecule, synthesis intermediate, catalyst, radiation sensitive compound, autocatalyst, and protecting group removing product. Applicants respectfully disagree with the Examiner's assertion because a person of ordinary skill in the art given the benefit of the present disclosure would understand the scope and meaning of such terms.

For example, the term "protecting group" is defined at page 10, lines 9-31 of the specification:

> 6. Protective Group: A material which may be selectively removed to expose an active site such as, in the specific example of an amino acid, an amine group. By way of illustration, protecting groups include but are not limited to those that are photolabile (see Fodor et al., PCT Publication No. WO 92/10092 (previously incorporated by reference), U.S. Ser. No. 07/971,181, filed Nov. 2, 1992, and U.S. Ser. No. 08/310,817, filed Sep. 22, 1994 (all of which are incorporated herein by reference in their entirety for all purposes)), acid labile, and base labile. For an extensive listing of protective groups useful in the practice of the present invention, see also Greene, T. W. and Wuts, P. G. M., Protective Groups in Organic Synthesis, (1991), incorporated herein by reference in its entirety for all purposes. Useful representative acid sensitive protective groups include dimethoxytrityl (DMT), tert-butylcarbamate (tBoc) and trifluoroacetyl (Tfa). Useful representative base sensitive protective groups include 9-fluorenylmethoxycarbonyl (Fmoc), isobutyrl (iBu), benzoyl (Bz) and phenoxyacetyl (pac). Other protective groups include acetamidomethyl, acetyl, tert-amyloxycarbonyl, benzyl, benzyloxycarbonyl, 2-(4-biphenylyl)-2-propyloxycarbonyl, 2-bromobenzyloxycarbonyl, tert-butyl, tert-butyloxycarbonyl, 1-carbobenzoxamido-2,2,2-trifluoroethyl, 2,6-dichlorobenzyl, 2-(3,5-dimethoxyphenyl)-2-propyloxycarbonyl, 2,4-dinitrophenyl, dithiasuccinyl, formyl, 4-methoxybenzenesulfonyl, 4-methoxybenzyl, 4-methylbenzyl, o-nitrophenylsulfenyl, 2-phenyl-2-propyloxycarbonyl, .alpha.-2,4,5-

> tetramethylbenzyloxycarbonyl, p-toluenesulfonyl, xanthenyl, benzyl ester, N-hydroxysuccinimide ester, p-nitrobenzyl ester, p-nitrophenyl ester, phenyl ester, p-nitrocarbonate, p-nitrobenzylcarbonate, trimethylsilyl and pentachlorophenyl ester and the like.

Certainly, Applicants demonstrated possession of protective groups at the time of filing in view of the above disclosure.

In another example, the term "linker molecule" is disclosed at page 15, line 31-page 16, line 2 to have certain structural features:

> The linker molecule contains an acid- or base -removable protecting group. Useful linker molecules are well known to those skilled in the art and representative examples include oligo ethers such as hexaethylene glycol, oligomers of nucleotides, esters, carbonates, amides and the like.

and at page 23, lines 19-22:

> The linker molecules may be, for example, aryl acetylene, ethylene glycol oligomers containing 2-10 monomers or more, diamines, diacids, amino acids, or combinations thereof.

The scope of the term "synthesis intermediate" is readily identifiable to a person of ordinary skill in the art given the benefit of this disclosure. In certain embodiments, synthesis intermediates are linker molecules, nucleotides, polynucleotides, amino acids, polypeptides, etc. See claims 4-8 of parent U.S. Pat. No. 6,083,697. Certainly, such compounds have readily identifiable chemical structures (e.g., an amino acid has the general structure: $H_2N$-CHR-COOH).

In another example, the terms "catalyst" and "autocatalyst" are defined at page 9, lines 22-31:

> 9. Catalyst: A catalyst is any material that is not consumed in a chemical reaction and that affects the rate of the reaction. Reactions that are affected by catalysts are termed catalytic reactions. Autocatalytic reactions are reactions in which at least one of the products is also a catalyst for the reaction. An autocatalyst is a material that undergoes a reaction that produces a product

> that is also a catalyst for that same reaction. Some autocatalytic reactions have a relatively slow rate of reaction at the initial stage but the reaction is accelerated as it proceeds as more catalytic product is accumulated. Where a substance or a combination of substances undergoes two or more simultaneous reactions that yield different products, the distribution of products could be influenced by the use of a catalyst that selectively accelerates one reaction relative to the other(s).

The Examiner asserts that catalysts (and autocatalysts) need to have a common structural feature in order for those of skill in the art to know the scope these terms. Applicants respectfully note that those of skill in the art understand that catalysts by their very nature are typically described by their function, namely the description provided above. Accordingly, those of skill in the art given the benefit of the present disclosure would readily understand the meaning and scope of the terms catalyst and autocatalyst.

In another example, the term "radiation sensitive compound" is disclosed as being a photosensitive compound, such as a photoactivated acid catalyst, photoactivated catalyst, etc. See claims 2, 13, and 14 of parent U.S. Pat. No. 6,083,697. Just using photoactivated acid catalysts as an example, Applicants disclose at page 17, lines 8-27:

> Preferred PAACs include the 2, 1, 4 diazonaphthoquinone sulfonic acid esters and the 2, 1, 5-diazonaphthoquinone sulfonic acid esters. Other useful PACs include the family of nitrobenzyl esters, and the s-triazine derivatives. Suitable s-triazine acid generators are disclosed, for example, in U.S. Pat. No. 4,189,323, incorporated herein by reference. Non-ionic PAACs including halogenated non-ionic, photoacid generating compounds such as, for example, 1,1-bis[p-chorophenyl]-2,2,2-trichloroethane (DDT); 1,1- bis[p-methoxyphenyl]-2,2,2-trichloroethane; 1,2,5,6,9,10-hexabromocyclododecane; 1,10-dibromodecane; 1,1-bis[p-chlorophenyl]-2,2-dichloroethane; 4,4dichloro-2-(trichloromethyl) benzhydrol (Kelthane); hexachlorodimethyl sulfone; 2-chloro-6-(trichloromethyl) pyridine; o,o-diethyl-o-(3,5,6-trichloro-2-pyridyl) phosphorothionate; 1,2,3,4,5,6 hexachlorocyclohexane; N(1,1-bis[p-chlorophenyl]-2,2,2 trichloroethyl)acetamide; tris [2,3-dibromopropyl]isocyanurate; 2,2-bis [p-chlorophenyl]-1,1 dichloroethylene; tris [trichloromethyl] striazine; and their isomers, analogs, homologs, and residual compounds are also suitable for some applications. Suitable PAACs are also disclosed in European Patent Application Nos. 0164248 and 0232972, both incorporated by reference for all purposes. PAACs that are particularly preferred for deep UV exposure include 1,1-bis (p-

> chlorophenyl)-2,2,2-trichloroethane (DDT); 1,1-bis (p-methoxyphenol)-2,2,2,-trichloroethane; 1,1-bis(chlorophenyl)-2,2,2 trichloroethanol; tris (1,2,3-methanesulfonyl) benzene; and tris (trichloromethyl) triazine.

Certainly, those of skill in the art given the benefit of the present disclosure can readily identify radiation sensitive compounds by their structural features.

Lastly, protecting group removing products are disclosed at page 23, lines 29-31 as being acids, in certain embodiments:

> Since in one embodiment of the present invention, the RAC catalyzes cleavage of the enhancer to produce an acid used to remove an acid-labile protective group, the effective quantum yield of the radiation is much larger than one, resulting in a high sensitivity.

Certainly, those of skill in the art given the benefit of the present disclosure can identify acids.

Accordingly, Applicants' description of the presently claimed methods in the specification described above demonstrates Applicants' possession of the claimed compounds at the time of filing.

## VI. Claims 1-3, 6, 12, 14-15, 52-54, 56-60, 70-71, and 73-76 Satisfy the Enablement Requirement

Claims 1-3, 6, 12, 14-15, 52-54, 56-60, 70-71, and 73-76 stand rejected under 35 U.S.C. § 112, ¶1 pursuant to the enablement requirement. Applicants respectfully traverse this rejection.

Applicants established above that the present disclosure adequately demonstrates that Applicants were in possession of the claimed subject matter (protecting groups, linker molecules, synthesis intermediates, catalysts, radiation sensitive compounds, autocatalysts, and protecting group removing products) at the time of filing. Likewise, through Applicants' teachings and disclosure a person of ordinary skill in the art would understand how to make and use such compounds based on the test set forth in In re Wands. 858 F.2d 731, 737 (Fed. Cir. 1988).

1. **Breadth of Claims/Nature of Invention**

The Examiner asserts that the specification does not provide distinguishing structural attributes for any of the above-referenced compounds, e.g., regarding the number of atoms, types of atoms, etc.

In response, for each of the above-referenced compounds, Applicants demonstrated above that the application at the time of filing provided certain structural features indicative of each of the compounds. For example, the application discloses at page 10, lines 9-31 (quoted above) many exemplary suitable protecting groups. Further, Applicants disclose, e.g., at page 15, line 31-page 16, line 2 that linker molecules have certain structural features, such as being "oligo ethers such as hexaethylene glycol, oligomers of nucleotides, esters, carbonates, amides and the like." Certainly, in view of the above discussion, Applicants provide in the specification distinguishing structural features for each of the claimed compounds. Accordingly, the specification as filed supports the breadth of each of the claimed compounds.

2. **State of the Prior Art/Level of Predictability**

The Examiner appears to assert at page 14 that since the applicable art is unpredictable, "a parallel placement method for use in the preparation of arrays of organic compounds is not presently available." The Examiner cites certain publications that supposedly indicate unfavorable research results.

In response, Applicants point out that unfavorable results such as "poor chemical quality" or "diminished yields" are not an indicator of the level of predictability in the applicable art. If anything, such results increase the level of predictability for the applicable art insofar as those of skill in the art viewing such references would then expect to obtain little success when performing similar experiments. Similarly, a publication supposedly disclosing that processes of

photochemical deprotection "never adequately produced useful peptide arrays" would lead one to expect to obtain little success when performing similar experiments. To the extent that Applicants achieve success in view of the state of the art, Applicants have achieved unexpected results.

3. **Level of Skill**

Applicants agree that the level of required skill in the applicable art is high.

4. **Amount of Direction/Existence of Working Examples**

The Examiner asserts that the application provides "only examples of 'acid' photoactivated catalysts used solely for polynucleotide synthesis."

In response, Applicants point out that the specification states that photoactivated catalysts are not necessarily acid catalysts. For example, the specification states at page 3, lines 12-14: "Such photosensitive compounds include what are generally referred to as radiation-activated catalysts (RACs), and more specifically photo activated catalysts (PACs)." The specification further defines RACs at page 12, lines 1-3: "A radiation activated catalyst (RAC) is a compound or group which produces at least one catalyst when exposed to radiation. RACs include but are not limited to radicals, acids, bases, ions, and metals." Certainly, the breadth of Applicants' photoactivated catalysts encompasses compounds other than acids.

5. **Quantity of Experimentation**

While the chemical arts are generally known to be inherently unpredictable, that does not mean that one would not be able to make the claimed compounds given the benefit of Applicants' disclosure and the state of the art. Besides the various examples disclosed throughout the specification, Applicants disclose Examples I-IV from pages 26-33, which disclose *inter alia* the removal of protecting groups by chemical amplification. Thus, the state of

the art is such that one of ordinary skill would be able to make and use the invention commensurate in scope with the cited claims.

6. **Level of Skill in the Art**

It is without a doubt that the level of skill in the pertinent art is high. The Examiner admits this point at page 7 of the Office Action.

Applicants have established through the above analysis with evidence that the cited compounds are enabled by Applicants' disclosure. Accordingly, removal of the rejection is requested at this time.

## VII. Claims 1-3, 6, 12, 14-15, 52-54, 56-60, 70-71 and 73-76 Stand Rejected under 35 U.S.C. § 112, ¶2.

Claims 1-3, 6, 12, 14-15, 52-54, 56-60, 70-71 and 73-76 stand rejected under 35 U.S.C. § 112, ¶2 as being indefinite. Applicants respectfully traverse this rejection.

The Examiner asserts that the terms radiation sensitive compound, autocatalytic compound, and protecting group removing product are indefinite as they supposedly are not defined with any chemical or physical characteristic, but only by functional properties.

In response, Applicants point out that each of the above-referenced terms is described in the specification by certain structural features. Applicants discuss this point in previous sections above. Notably, Applicants discuss radiation sensitive compounds and photosensitive compounds, such as a photoactivated acid catalyst, photoactivated catalyst, etc. See claims 2, 13, and 14 of parent U.S. Pat. No. 6,083,697. Just using photoactivated acid catalysts as an example, Applicants disclose various chemical compounds, which are defined by their structures, at page 17, lines 8-27 (quoted above). Likewise, autocatalytic compounds, which are defined at page 9,

lines 22-31 of the specification, are discussed above. Further, protecting group removing products are disclosed at page 23, lines 29-31.

Lastly, the Examiner asserts that the term catalyst system is not definite. However, the specification discloses at page 3, lines 27-28: "The photo activated catalyst by itself or in combination with additional catalytic components is referred to herein as a catalyst system." Applicants teach throughout the specification certain catalytic components, which include, but are not limited to, autocatalysts, catalysts, etc.

Accordingly, the meaning of each of the cited terms is definite.

## VIII. Claims 1-2, 6, 12, 15, 52-54, 56-60, 70-71, and 73-76 Are Patentable over Pirrung

Claims 1-2, 6, 12, 15, 52-54, 56-60, 70-71, and 73-76 stand rejected under 35 U.S.C. § 102(b) over Pirrung. Applicants respectfully traverse this rejection.

The cited claims are patentable over Pirrung at least because Pirrung fails to disclose an autocatalytic compound, as defined all pending independent claims. The Examiner asserts that the NVOC group disclosed by Pirrung is a photoremovable protecting group as well as an autocatalytic compound. However, this is simply not the case. Applicants discuss above with respect to Holmes, that the NVOC group is not a material that undergoes a reaction that produces a product that is also a catalyst for that same reaction, i.e., is an autocatalyst. Indeed, Pirrung, like Holmes, does not disclose that the NVOC group catalyzes anything, let alone a reaction for the formation of more NVOC. In that regard, all cited claims are patentable over Pirrung at least because Pirrung does not disclose a protecting group removing product generated by an autocatalytic compound.

Further, the Examiner cites Pirrung at column 28, line 12-66 for the synthesis of YGGFL. It is worth noting that Pirrung's use of t-BOC to protect the aminopropyl group as seen in FIG. 14A, is not a disclosure of an autocatalytic compound. As discussed above with respect to MacDonald, t-BOC is a protecting group and not an autocatalyst. That is, t-BOC group in MacDonald does not undergo a reaction that produces a product that is also a catalyst for that same reaction.

Similarly, independent claims 1 and 52 and all claims dependent therefrom are patentable over Pirrung because Pirrung does not disclose either explicitly or inherently a catalyst produced by a radiation sensitive compound or a photosensitive acid compound, respectively. Accordingly, removal of the present rejection is respectfully requested at this time.

## IX. Claims 1-3, 6, 12, 14-15, 52-54, 56-60, 70-71, and 73-76 Are Patentable over the '721 Publication

Claims 1-3, 6, 12, 14-15, 52-54, 56-60, 70-71, and 73-76 stand rejected under the obviousness-type double patenting doctrine as to the '721 publication. Applicants are submitting herewith a terminal disclaimer, which, as the Examiner indicates at page 19 of the Office Action, can be used to overcome an obviousness-type double patenting rejection. As such, the present rejection is overcome.

## X. Claims 1-3, 6, 12, 14-15, 52-54, 56-60, 70-71, and 73-76 Are Patentable over the '697 Patent

Claims 1-3, 6, 12, 14-15, 52-54, 56-60, 70-71, and 73-76 stand rejected under the obviousness-type double patenting doctrine as to the '697 patent. Applicants are submitting herewith a terminal disclaimer, which, as the Examiner indicates at page 19 of the Office Action,

can be used to overcome an obviousness-type double patenting rejection. As such, the present rejection is overcome.

## XI. Conclusion

Applicants have shown that Holmes, MacDonald, and Pirrung each do not anticipate the pending claims. Further, Applicants have clearly demonstrated that the specification as filed meets the written description and enablement requirements of § 112, ¶1 and the definiteness requirement of § 112, ¶2.

Having addressed all outstanding issues, Applicants respectfully request allowance of the case at this time. To the extent the Examiner believes that it would facilitate allowance of the case, the Examiner is invited to telephone the undersigned at the number below.

Respectfully submitted,

Dated: November 24, 2003

John P. Iwanicki, Reg. No. 34,628
BANNER & WITCOFF, LTD.
28 State Street, 28th Floor
Boston, MA 02109
(617) 720-9600

ORGANIC LETTERS
2002
Vol. 4, No. 14
2293–2296

# Synthesis and Applications of End-Labeled Neoglycopolymers

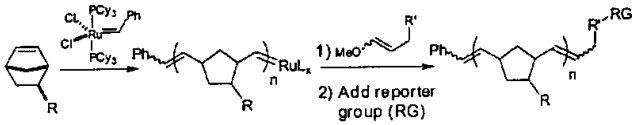
**Robert M. Owen,[†] Jason E. Gestwicki,[‡] Travis Young,[†] and Laura L. Kiessling*[,†,‡]**

*Departments of Chemistry and Biochemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706*

*kiessling@chem.wisc.edu*

**Received March 24, 2002**

## ABSTRACT




**Neoglycopolymers that vary in length and contain a single fluorescent reporter group were synthesized using ring-opening metathesis polymerization (ROMP). The utility of these materials is demonstrated by the development of a cellular binding assay for L-selectin, a cell surface protein that plays a role in inflammation. The data reveal that these multivalent ligands interact with multiple copies of L-selectin.**

Multivalent binding events are ubiquitous in biological systems. For example, multivalent protein–carbohydrate complexation events are important in cell adhesion, host–pathogen interactions, and the immune response.[1] Despite the importance of multivalent binding, mechanistic information is often lacking. One critical issue that arises when a multivalent ligand interacts with a cell-surface protein is whether it binds multiple copies of the target receptor. Though this binding mode is often invoked with multivalent ligands, it is difficult to determine the stoichiometry of these receptor–ligand complexes.[2] We sought to address this issue by examining multivalent ligand binding to cell-surface L-selectin.

L-Selectin is displayed on the cell-surface of leukocytes where it participates in the recruitment of these cells to the inflamed endothelium.[3,4] The natural ligands for L-selectin are multivalent; they display multiple copies of sulfated carbohydrates.[4] L-Selectin is localized to regions on the cell-surface,[5] and dimerization of L-selectin enhances its ability to bind to endothelial cells.[6] These data implicate multivalency in the interaction of L-selectin and its ligands at the cell surface, but direct evidence has been difficult to obtain.

Synthetic multivalent displays have emerged as powerful tools for investigating multivalent binding.[7] Polymeric ligands that incorporate both binding elements and a reporter group are especially useful for visualizing and quantitating binding interactions.[8] A reporter moiety, such as a fluorescent or biotin group, is typically introduced through the polymerization of an appropriately substituted monomer or through

[†] Department of Chemistry.
[‡] Department of Biochemistry.

(1) (a) Lee, Y. C.; Lee, R. T. *Acc. Chem. Res.* **1995**, *28*, 321–327. (b) Kiessling, L. L.; Pohl, N. L. *Chem. Biol.* **1996**, *3*, 71–77. (c) Roy, R. *Curr. Opin. Struct. Biol.* **1996**, *6*, 692–702. (d) Mammen, M.; Choi, S.-K.; Whitesides, G. M. *Angew. Chem., Int. Ed.* **1998**, *37*, 2754–2794. (e) Feizi, T. *Immunol. Rev.* **2000**, *173*, 79–88. (f) Sacchettini, J. C.; Baum, L. G.; Brewer, C. F. *Biochemistry* **2001**, *40*, 3009–3015. (g) Bertozzi, C. R.; Kiessling, L. L. *Science* **2001**, *291*, 2357–2364.

(2) (a) Gestwicki, J. E.; Strong, L. E.; Kiessling, L. L. *Angew. Chem., Int. Ed.* **2000**, *39*, 4567–4570. (b) Cairo, C. W.; Gestwicki, J. E.; Kanai, M.; Kiessling, L. L. *J. Am. Chem. Soc.* **2002**, *124*, 1615–1619.

(3) (a) Springer, T. A. *Cell* **1994**, *76*, 301–314. (b) Steeber, D. A.; Tedder, T. F. *Immunol. Res.* **2000**, *22*, 299–317.

(4) (a) Yeh, J. C.; Hiraoka, N.; Petryniak, B.; Nakayama, J.; Ellies, L. G.; Rabuka, D.; Hindsgaul, O.; Marth, J. D.; Lowe, J. B.; Fukuda, M. *Cell* **2001**, *105*, 957–969. (b) Hemmerich, S.; Rosen, S. D. *Glycobiology* **2000**, *10*, 849–856. (c) Yarema, K. J.; Bertozzi, C. R. *Curr. Opin. Chem. Biol.* **1998**, *2*, 49–61.

(5) Bruehl, R. E.; Springer, T. A.; Bainton, D. F. *J. Histochem. Cytochem.* **1996**, *44*, 835–844.

(6) (a) Galustian, C.; Childs, R. A.; Yuen, C. T.; Hasegawa, A.; Kiso, M.; Lubineau, A.; Shaw, G.; Feizi, T. *Biochemistry* **1997**, *36*, 5260–5266. (b) Li, X.; Steeber, D. A.; Tang, M. L. K.; Farrar, M. A.; Perlmutter, R. M.; Tedder, T. F. *J. Exp. Med.* **1998**, *188*, 1385–1390.

(7) (a) Liang, R.; Yan, L.; Loebach, J.; Ge, M.; Uozumi, Y.; Sekanina, K.; Horan, N.; Gildersleeve, J.; Thompson, C.; Smith, A.; Biswas, K.; Still, W. C.; Kahne, D. *Science* **1996**, *274*, 1520–1522. (b) Kiessling, L. L.; Strong, L. E. *Top. Organomet. Chem* **1998**, *1*, 199–231. (c) Lundquist, J. J.; Toone, E. J. *Chem. Rev.* **2002**, *102*, 555–578. (d) Houseman, B. T.; Mrksich, M. *Top. Curr. Chem.* **2002**, *218*, 1–44. (e) Lindhorst, T. K. *Top. Curr. Chem.* **2002**, *218*, 201–235.

(8) Reviewed in: Bovin, N. V. *Glycoconjugate J.* **1998**, *15*, 431–446.

10.1021/ol0259239 CCC: $22.00 
Published on Web 06/11/2002

random conjugation to the polymer backbone.[9] Although useful for many purposes, these strategies provide limited control over the number and location of the introduced functionality. Polymers possessing a single end-label, however, can be selectively immobilized on surfaces to provide novel materials for physical[10] or biological studies.[11] We envisioned that multivalent ligands varying in valency but containing a single fluorophore per ligand could provide unique mechanistic information. If multivalent binding occurs, saturating concentrations of ligands differing only in their valency would give rise to different cell surface fluorescence intensities (Figure 1). In contrast, monovalent

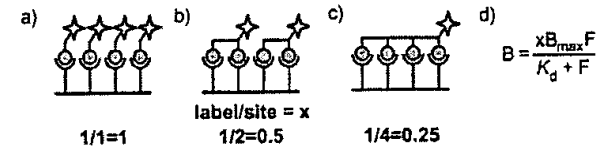



**Figure 1.** End-labeled materials can be used to reveal multivalent interactions. The number of labels bound depends on the number of binding sites occupied per multivalent ligand (parts a–c). The binding of end-labeled multivalent ligands at saturation ($B_{max}$) will vary according to valency (equation shown in part 1d, $B$ is moles of bound ligand, $F$ is the free multivalent ligand concentration).

binding interactions would yield identical fluorescence intensities at saturation.

ROMP[12] provides a means to generate end-labeled materials varying in valency. When ROMP is living,[13] an active carbene (e.g., **3**) remains at the polymer terminus after consumption of the monomer (Figure 2). The metal alky-

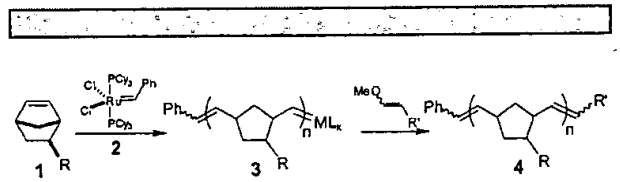


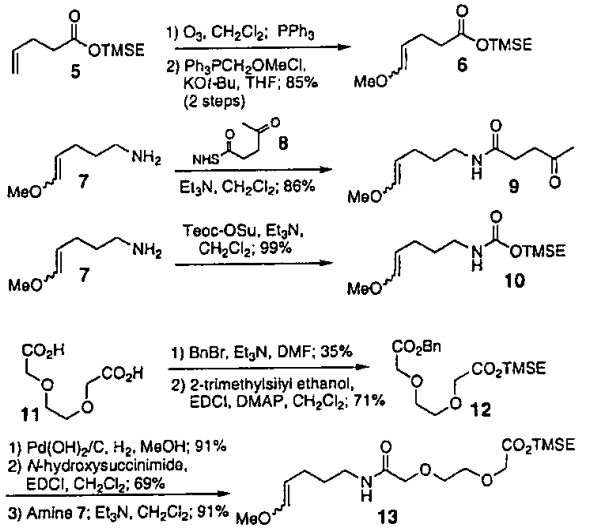

**Figure 2.** Termination of ruthenium carbene-initiated ROMP reactions with an enol-ether provides an end-labeled material (**4**).

lidene **3** can undergo further transformations. For example, Schrock and co-workers treated a polymer possessing a terminal molybdenum carbene with aldehydes to provide materials with unique redox and luminescence properties.[10] The highly functional group tolerant ruthenium initiator **2** developed by the Grubbs group also has been used to synthesize end-labeled polymers. Ruthenium alkylidene-initiated reactions conducted in the presence of chain-transfer agents provide symmetrically terminated materials.[14] Alternatively, a terminal aldehyde can be installed by treatment of **3** with molecular oxygen.[15] We have shown that a substituted enol ether can be used to produce specifically, end-labeled polymers, such as **4**.[16]

Here, we demonstrate the effectiveness of this strategy for the synthesis of labeled polymers varying in length and terminal functionality.

**Termination Efficiency.** Because our objective was to generate multivalent ligands of different lengths, the capping process must be efficient. To identify effective termination agents, enol ether derivatives possessing unique functional groups, including masked acid (**6**, **13**), ketone (**9**), and masked amine (**10**) groups, were synthesized (Scheme 1).

**Scheme 1.** Synthetic Routes to Acid-, Ketone-, and Amine-Substituted Capping Agents




The functional groups chosen provide the means to install reporter groups under mild conditions using commercially available reagents. The capping agents were assembled from simple building blocks (Scheme 1). Enol ether **6** was obtained in excellent yield from the known 2-trimethylsilyl ethyl (TMSE) ester (**5**) of 4-pentenoic acid.[17] Ozonolysis of **5** followed by Wittig reaction yielded the desired vinyl ether **6** as a 1:3 *cis:trans* mixture. Although *cis*-enol ethers have been shown to react more quickly with ruthenium carbene

(9) (a) Ticha, M.; Kocourek, J. *Carbohydr. Res.* **1991**, *213*, 339–343. (b) Kamitakahara, H.; Suzuki, T.; Nishigori, N.; Suzuki, Y.; Kanie, O.; Wong, C. H. *Angew. Chem., Int. Ed.* **1998**, *37*, 1524–1528.

(10) (a) Albagli, D.; Bazan, G. C.; Schrock, R. R.; Wrighton, M. S. *J. Am. Chem. Soc.* **1993**, *115*, 7328–7334. (b) Albagli, D.; Bazan, G. C.; Schrock, R. R.; Wrighton, M. S. *J. Phys. Chem.* **1993**, *97*, 10211–10216.

(11) Gestwicki, J. E.; Cairo, C. W.; Mann, D. A.; Owen, R. M.; Kiessling, L. L. *Anal. Biochem.*, in press.

(12) Trnka, T. M.; Grubbs, R. H. *Acc. Chem. Res.* **2001**, *34*, 18–29.

(13) Ivin, K. J.; Mol J. C. *Olefin Metathesis and Metathesis Polymerization*; Academic Press: London, 1997.

(14) (a) Gibson, V. C.; Okada, T. *Macromolecules* **2000**, *33*, 655–656. (b) Morita, T.; Maughon, B. R.; Bielawski, C. W.; Grubbs, R. H. *Macromolecules* **2000**, *33*, 6621–6623.

(15) (a) Gordon, E. J. Ph.D. Dissertation, University of Wisconsin, Madison, WI, 1998. (b) Biagini, S. C. G.; Davies, R. G.; Gibson, V. C.; Giles, M. R.; Marshall, E. L.; North, M. *Polymer* **2001**, *42*, 6669–6671.

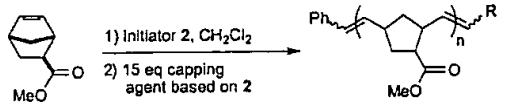
(16) Gordon, E. J.; Gestwicki, J. E.; Strong, L. E.; Kiessling, L. L. *Chem. Biol.* **2000**, *7*, 9–16.

(17) Krafft, M. E.; Cheung, Y. Y.; Wright, C.; Cali, R. *J. Org. Chem.* **1996**, *61*, 3912–3915.

2,[18] the isomers were not separated because an excess of the mixture (15 equiv) was used in the termination step (vide infra). Capping agents **9**, **10**, and **13**, were all generated from known amine **7**.[19] The ketone-containing capping agent **9** could be synthesized in a single step from **7** and the *N*-hydroxysuccinimide (NHS) ester of levulinic acid (**8**).[20] Similarly, compound **10** could be accessed from the reaction of **7** with 2-trimethylsilylethyl *N*-succinimidyl carbonate (Teoc-OSu).[21] Preparation of the ethylene glycol-based capping agent (**13**) required desymmetrization of the diacid **11** to give the diester **12**. The benzyl ester was converted to the NHS ester via the acid. The enol ether was installed under standard amide bond forming conditions to afford capping agent **13**. The efficiencies of **6**, **9**, **10**, and **13** in terminating ROMP reactions were then assessed.

To test the effectiveness of our capping strategy, monomer **14** was polymerized using ruthenium initiator **2**, and the reaction was terminated by the addition of a capping agent. Methyl ester **14** was used as a monomer because the resulting polymer (**15**) could be readily separated from excess capping agent by silica gel chromatography. Each of the capping agents was used in a 15-fold excess as a mixture of *cis*:*trans* isomers. The termination reaction was monitored by $^1$H NMR spectroscopy. Complete consumption of the propagating carbene by each terminating agent occurred within 3 h. After purification of the resulting polymers, the capping efficiencies were determined by comparing the integration of the signal due to the polymer phenyl protons relative to a diagnostic signal from the capping agent (the trimethylsilyl group for compounds **6**, **10**, and **13** or the protons α to the ketone for **9**). The termination efficiencies were excellent for **6** and **10**, good for **9**, and moderate for compound **13** (Table 1). The lower efficiencies for **9** and **13** may arise from their amide groups, which may chelate to the ruthenium alkylidene species.[22] The termination reaction of the carbamate-containing capping agent **10**, which is less Lewis basic than an amide,[23] is more efficient.

**Table 1.** Capping Efficiencies for the Polymerization of Alkene **14**



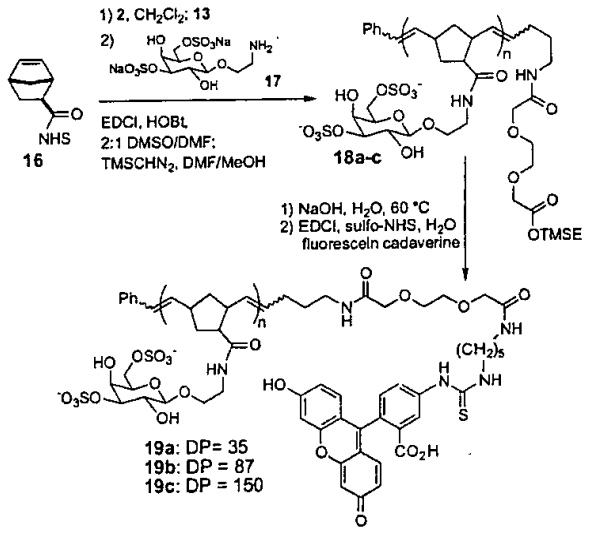

**14** **15**

| entry | capping agent | capping efficiency[a] |
|---|---|---|
| 1 | **6** | >95%/93% |
| 2 | **9** | 80%/86% |
| 3 | **10** | >95%/92% |
| 4 | **13** | 64%/68% |

[a] Efficiencies are an average of a minimum of two separate experiments and were determined for monomer-to-catalyst ratios of 15:1/50:1.

**End-Labeled Polymers as Cell-Surface Probes.** Given that end-labeled polymers could be generated efficiently, these materials were used to investigate ligand binding to L-selectin-displaying cells. We found previously that multivalent ligands presenting 3,6-disulfogalactose residues bind to L-selectin.[16,24] We envisioned that fluorescent derivatives could be used to assay binding[16,25] and to assess the importance of multivalent interactions. A series of polymers in which the degree of polymerization was varied was generated by controlling the ratio of monomer **16**[26] to ruthenium alkylidene initiator **2**. The polymerization reactions were terminated by the addition of **13** to afford electrophilic polymers that could be sequentially functionalized. Treatment with the amine-bearing 3,6-disulfogalactose derivative (**17**)[27] yielded the desired polymers (**18a**–**c**) (Scheme 2). Hydrolysis of the terminal ester and conjugation of fluorescein cadaverine to the resulting amine group provided the target labeled ligands (**19a**–**c**).[16] The interaction of these materials with Jurkat cells displaying L-selectin was then assessed.

**Scheme 2.** Synthesis of Labeled 3,6-Disulfogalactose-substituted Polymers **20a**–**c** for L-Selectin Binding Assay. The Degree of Polymerization (DP[30]) Was Determined from the $^1$H NMR Spectra




Jurkat cells were exposed to polymers **19a**–**c** or a fluorescently labeled anti-L-selectin antibody (Figure 3A)

(18) Hillmyer, M. A.; Grubbs, R. H. *Macromolecules* **1995**, *28*, 8662–8667.
(19) Wasserman, H. H.; Cook, J. D.; Vu, C. B. *J. Org. Chem.* **1990**, *55*, 1701–1702. Typical *cis*:*trans* ratios obtained ranged from 1:2 to 1:4.
(20) Lee, J. H.; Baker, T. J.; Mahal, L. K.; Zabner, J.; Bertozzi, C. R.; Wiemer, D. F.; Welsh, M. J. *J. Biol. Chem.* **1999**, *274*, 21878–21884.
(21) Schute, R. E.; Rich, D. H. *Synthesis* **1987**, *4*, 346–349.
(22) Choi, T. L.; Chatterjee, A. K.; Grubbs, R. H. *Angew. Chem., Int. Ed.* **2001**, *40*, 1277–1279.
(23) Grana, A. M.; Mosquera, R. A. *J. Comput. Chem.* **1999**, *20*, 1444–1454.
(24) Gordon, E. J.; Strong, L. E.; Kiessling, L. L. *Bioorg. Med. Chem.* **1998**, *6*, 1–7.
(25) Wittmann, V.; Datta, A. K.; Koeller, K. M.; Wong, C. H. *Chem. Eur. J.* **2000**, *6*, 162–171.
(26) Strong, L. E.; Kiessling, L. L. *J. Am. Chem. Soc.* **1999**, *121*, 6193–6196.
(27) Manning, D. D.; Hu, X.; Beck, P.; Kiessling, L. L. *J. Am. Chem. Soc.* **1997**, *119*, 3161–3162.

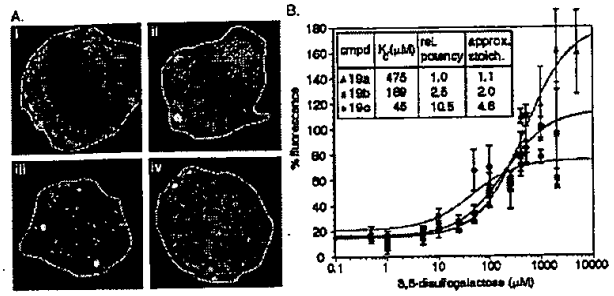



**Figure 3.** (A) Treatment of Jurkat cells with (i) fluorophore-labeled anti-L-selectin antibody, (ii) polymer **19a**, (iii) **19b**, or (iv) **19c**. (B) Results from the L-selectin binding assay. The binding of **19a**−**c** is reversed upon addition of unlabeled 3,6-disulfogalactose or sialyl Lewis x-bearing polymers (data not shown). Concentrations refer to 3,6-disulfogalactose residue concentration. The approximate stoichiometry is defined as the number of copies of L-selectin that interact simultaneously with a polymer of a given DP.[28]

toscreen for ligand binding. As anticipated,[16] fluorescence microscopy experiments indicate that **19a**−**c** interact with Jurkat cells displaying L-selectin. To quantify the amount of bound ligand, Jurkat cells were treated with **19a**−**c** and washed, and the intensity of the fluorescence emission was assessed (Figure 3B). The resulting data were fit to the equation in Figure 1d to determine the dissociation constant ($K_d$) for each compound.[28] The $K_d$ values for the polymers[29] were dependent on the polymer length. Polymer **19c** (degree of polymerization,[30] DP of 150) had a potency that was ca. 10-fold greater than the shortest polymer (**19a**, DP of 35) on a saccharide residue basis. Although useful for comparison, this calculation underestimates the relative increase in potency. The individual saccharide (3,6-disulfogalactose) does not bind to L-selectin—a multivalent display is required.

Because the polymers possess a single fluorophore label, their ability to bind multiple copies of L-selectin at the cell surface could be determined. The concentration of polymer required to saturate the fluorescence intensity ($B_{max}$) was measured. Dividing the average amount of L-selectin on a Jurkat cell-surface by the calculated concentrations of bound ligand provides an estimate of the relative stoichiometry of the L-selectin−ligand complexes. By this analysis, polymer **19c** (DP of 150) bound approximately five copies of L-selectin while polymer **19a** (DP of 35) bound only one or two copies of L-selectin. These data indicate that compounds **19a**−**c** engage in multivalent interactions with L-selectin and that the stoichiometry of the resulting complex depends on the valency of the ligand. This result is consistent with studies demonstrating that neoglycopolymers can cluster proteins in solution and in the cell.[2,31] Compounds with more complex binding epitopes[32] may more effectively interact with multiple copies of cell-surface L-selectin.

Our previous studies indicate that multivalent but not monovalent ligands induce L-selectin shedding from white blood cells.[24] One explanation for this difference is that polymeric ligands cluster multiple copies of L-selectin but monovalent ligands do not. Because the ligands used in these studies induce L-selectin downregulation,[24] our results suggest that the ability of these ligands to modulate cell-surface levels of L-selectin depends on clustering.[33] This finding offers new opportunities for the design of L-selectin antagonists.[34] Our results also suggest mechanisms by which natural multivalent ligands may bind L-selectin.

In conclusion, end-labeled polymers are valuable probes of multivalent ligand−receptor interactions. Our data reveal that multivalent ligands varying in length with single end-labels can be generated using ROMP. The investigations described here highlight the utility of such materials for probing the mechanisms underlying multivalent binding at the cell surface. The ligands we have generated provide the means to investigate mechanistic aspects of cell-surface receptor−ligand interactions that have been inaccessible. We anticipate that the strategies outlined can be used to synthesize polymers that will be useful in a wide range of materials science and in vitro and in vivo biological applications.

**Acknowledgment.** This work was supported in part by the NIH (GM 55984). The UW-Madison Chemistry NMR facility is supported in part by NSF (CHE9629688 and CHE9208463) and the NIH (RR08389). R.M.O thanks Pharmacia Corporation and Eastman Chemical Company for fellowships. J.E.G. was supported by the NIH Biotechnology Training Program (GM08349). The authors thank R. J. Hinklin, C. W. Cairo, and J. K. Pontrello for helpful discussions and the laboratories of Profs S. Bednarek and P. Friesen (UW−Madison) for generous use of equipment.

**Supporting Information Available:** Experimental procedures and characterization data for compounds **6**, **9**, **10**, **12**, **13**, **15**, **18**, and **19**. This material is available free of charge via the Internet at http://pubs.acs.org.

OL0259239

(28) See the Supporting Information.
(29) $K_d$ values were calculated on a saccharide residue basis.
(30) Degree of polymerization (DP) was determined from the integration of $^1$H NMR signals in the phenyl region against those in the alkene region.

(31) (a) Burke, S. D.; Zhao, Q.; Schuster, M. C.; Kiessling, L. L. *J. Am. Chem. Soc.* **2000**, *122*, 4518−4519. (b) Gestwicki, J. E.; Strong, L. E.; Kiessling, L. L. *Chem. Biol.* **2000**, *7*, 583−591. (c) Gestwicki, J. E.; Strong, L. E.; Cairo, C. W.; Boehm, F. J.; Kiessling, L. L. *Chem. Biol.* **2002**, *9*, 163−169. (d) Gestwicki, J. E.; Kiessling, L. L. *Nature* **2002**, *415*, 81−84.
(32) (a) Sanders, W. J.; Gordon, E. J.; Dwir, O.; Beck, P. J.; Alon, R.; Kiessling, L. L. *J. Biol. Chem.* **1999**, *274*, 5271−5278. (b) Manning, D. D.; Bertozzi, C. R.; Pohl, N. L.; Rosen, S. D.; Kiessling, L. L. *J. Org. Chem.* **1995**, *60*, 6254−6255.
(33) Gordon, E. J.; Sanders, W. J.; Kiessling, L. L. *Nature* **1998**, *392*, 30−31.
(34) Simanek, E. E.; McGarvey, G. J.; Jablonowski, J. A.; Wong, C. H. *Chem. Rev.* **1998**, *98*, 833−862.

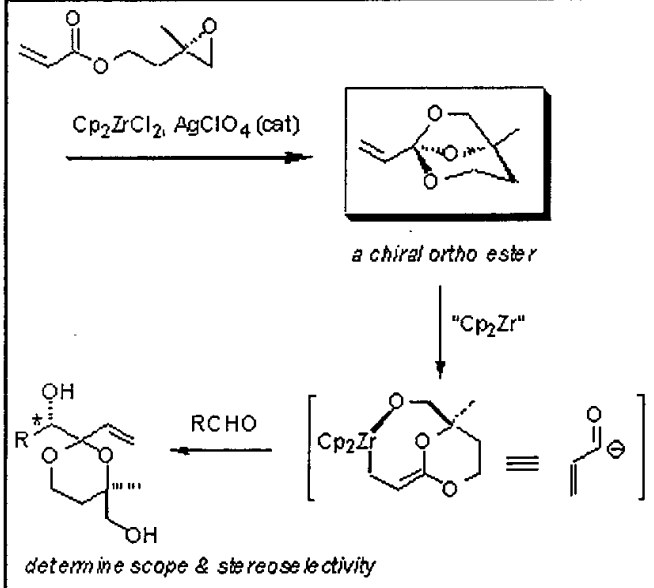
Research Experiences for Undergraduates — Measuring Molecules — Summer 2001-2003

# Research Projects, Summer 2002

The University of Pittsburgh's Department of Chemistry will conduct a Research Experiences for Undergraduates (REU) program, in Summer, 2002, the second year of a three year program. The projects available each year will vary and will range in scope from traditional to instrument intensive, from one discipline to multi-discipline in flavor, from specific target oriented to more open ended projects. The projects that are currently available are described here, in reverse alphabetical order according to the Professor's Last Name. Additional projects will be posted in the near future.

## Professor Peter Wipf

### *Measuring Asymmetric Inductions in Novel Synthetic Applications of Chiral Ortho Esters*




Ortho esters represent a class of masked acid derivatives that greatly modify the reactivity pattern of the parent carboxylates; however, they have found surprisingly little use in organic synthesis to date, and much of their chemistry remains to be explored. We are most interested in the use of ortho esters for new carbon-carbon bond formations, and we have developed a novel approach toward chiral ortho esters using a cationic zirconocene-catalyzed rearrangement reaction (Wipf, P.; Aslan, D. C., "Synthesis of bicyclic ortho esters by epoxy ester rearrangements and study of their ring opening reactions." *J. Org. Chem.* **2001**, *66*, 337-343). In the presence of zirconocene(II) complex, these chiral ortho esters can lead to the chiral g,g-dialkoxyallylic zirconium species **1** that is expected to provide asymmetric induction in the addition to aldehydes. The study of the scope of this reaction and the analysis of asymmetric induction by chiral HPLC is the topic of this project. [1/02]

## Professor Steve Weber

### *Using Gas-Phase Mass Spectrometric Measurement Methods to Understand and Design Solution Equilibria*

Are solution phase equilibria reflected in the gas phase? This important question has been raised by mass spectroscopists using electrospray who have observed peaks identified as adducts of the appropriate stoichiometry. We propose to build a remarkably simple apparatus in which the concentrations of adduct-forming species can be varied in a controlled, reproducible way. The device relies on Taylor-Aris dispersion in an open channel to dilute in a mathematically known way the concentrations of a pair of adduct-forming species. The student will modify (~1 week) the front end of the existing, custom-built mass spectrometer in order to simultaneously achieve conditions yielding Taylor-Aris behavior and efficient electrospray ionization. The student will then measure how concentrations of each of the relevant solutes changes in time following its injection into the flow stream both theoretically and using 'standard' (e.g. absorbance) detection. Finally, ESI-MS of the flow will be performed to determine how well correlated observations of adducts are with predictions based on known concentration profiles. [2/02]

### *Screening Measurements for Selective Receptor Design*

## Professor Steve Weber

How selective can an extraction be? Product purification, environmental cleanup, hemo-dialysis, analytical sample separation all rely upon extraction to achieve a partial purification. It would be beneficial to have selective extraction media to carry out these tasks. The key is to bury a selective receptor for the target molecule(s) in a very poor solvent so that only the molecule of choice is extracted. The student will work with a library of receptors allowing the student to choose the target. Once a receptor has been chosen by a screening procedure it will be chemically modified to make it compatible with a fluorocarbon solvent. The solution of the suitably modified receptor in a fluorocarbon solvent will be examined by physical methods (e.g. 1H-NMR), and applied to selective extraction.3. Peptide-copper complexes: Structure and reactivity. Cu(II)-peptide complexes are important both as tools for analyses and as biological entities. How does the peptide structure influence the reaction rate, stability, and redox potential of the complex? Several physical methods will be used, including $^{1}H$, $^{13}C$-NMR (there is some broadening, but the spectra are useful), UV-vis, rotating ring disc electrode. The results can be applied to the important field of proteomics as a method for the quantitative determination of proteins and peptides. [2/02]

## Professor Steve Weber

***Measurements to Characterize Peptide-Copper Complexes***

Peptide-copper complexes: Structure and reactivity. Cu(II)-peptide complexes are important both as tools for analyses and as biological entities. How does the peptide structure influence the reaction rate, stability, and redox potential of the complex? Several physical methods will be used, including $^{1}H$, $^{13}C$-NMR (there is some broadening, but the spectra are useful), UV-vis, rotating ring disc electrode. The results can be applied to the important field of proteomics as a method for the quantitative determination of proteins and peptides. [2/02]

## Professor Gilbert Walker

***Controlling and Measuring the Release of Nitric Oxide in Bloodstream Implants.***

Nitric oxide is a principal signaling agent for neurotransmission, cell death, and thrombosis. How does NO get around so much? What are its protein targets and how do they change conformation to reflect NO binding? Can polymer substrates that release NO be prepared for biomedical implants, such as artificial capillaries? How do cells respond to this invasion? Work on these and related problems in a cross-disciplinary team that includes cell biologists, chemists and chemical engineers. Your job would be to measure the rate of nitric oxide release from specialized polymer surfaces into environments that mimic the bloodstream. [1/02]

## Professor Gilbert Walker

***Single Molecule Measurements of Protein Conformations at Artificial Implant Surfaces.***

Protein and polymer folding at surfaces is important for signaling between cells and for adhesion at surfaces. Identifying these folding states and relating them to specific functions is a holy grail of biomedical science. We have developed a variety of single molecule scanning probe techniques to identify surface folding states, including surface force and spectroscopic methods. This summer project involve identifying the different folding states of the adhesion protein fibronectin, which is being used to seed implant surfaces for endothelial cell adhesion. You would use single molecule force spectroscopic techniques, which allow you to unfold proteins, one domain at a time. This work is working towards minimization of thrombosis and related aspects of cardiovascular disease. [1/02]

## Professor David Waldeck

***Measuring Electron Tunneling in Organic Molecules.***

Electron transfer is a primary step in many important chemical and biological transformations. Electron transfer between molecular residues is likely to play a central role in future molecular devices. Our research program is studying the primary interactions between molecular subunits that control electron transfer reactions. A particular example of one such system is illustrated by the Donor-Bridge-Acceptor molecule in the

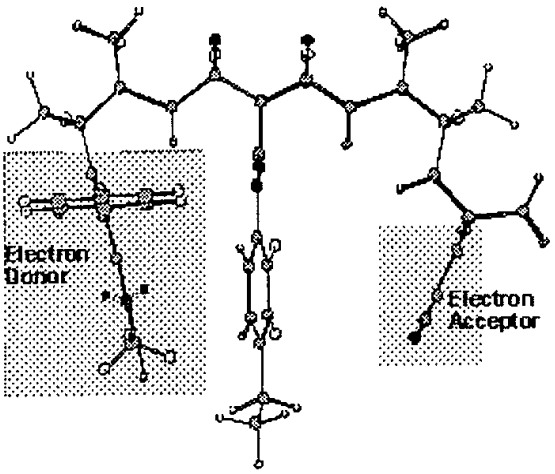
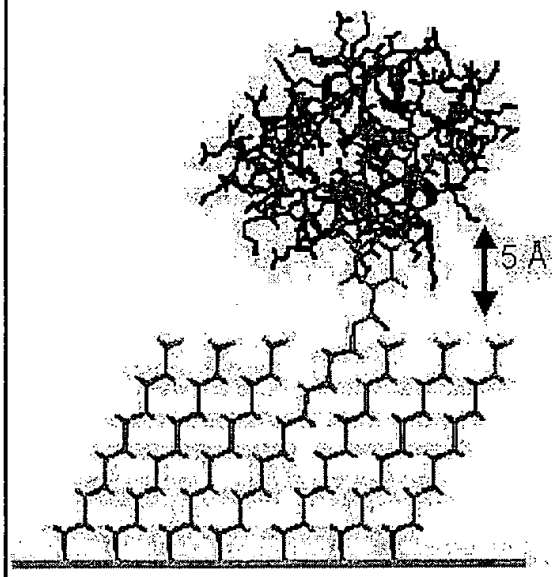



diagram. Recently we demonstrated that the electron tunnels from the naphthalene donor moiety through the phenyl ring of the bridge to be captured on the electron acceptor and that the phenyl ring should lie in the 'line-of-sight' between the donor and acceptor (Ref 1). Future studies will examine the role played by the electronic structure of the phenyl moiety and its motion on the electron transfer reaction. [1/02]

1. A M.. Napper, I. Read, D. H. Waldeck, N. J. Head, A. M. Oliver and M. N. Paddon-Row, *AAn Unequivocal Demonstration of the Importance of Nonbonded Contacts in the Electronic Coupling between Electron Donor and Acceptor Units of Donor-Bridge-Acceptor Molecules, J. Am. Chem. Soc., 122* (2000) 5220-5221.

## Professor David Waldeck

### *Measurements Aimed at Understanding Electron Transfer Between Electrodes and Biomolecules.*

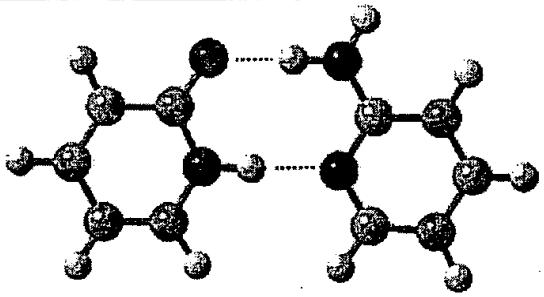



Electron transfer at electrode interfaces is of immense importance to a variety of current and future technologies. Interfacial electron transfer is central to processes that range from the ancient (corrosion is an excellent example of an old unsolved problem with large monetary implications) to the future (e.g., molecular electronics and bioelectronics). Our group is investigating the use of organic molecular films (SAMs) to control the electron transfer between an electrode and a surface bound biomolecule. The diagram illustrates the underlying architecture used in this approach – a metal electrode is coated with a monolayer thick organic film that can selectively bind a biomolecule. This architectural control allows us to investigate the electron transfer kinetics and how it depends on properties of the system, such as chain length, composition, dipole moment and others. [1/02]

1. H. Yamamoto, H. Liu, and D. H Waldeck "Immobilization of Cytochrome C at Au Electrodes by Ligation between a Pyridine Terminated SAM and the Heme of Cytochrome" *Chem. Commun.* 2001, 1032.

## Professor David Pratt

### *Measuring Hydrogen Bond Lengths in the Gas Phase.*



The structure of this dimer mimics the two hydrogen bonds in the AT base pair. Pitt undergraduate Rob Roscioli working with Prof. Pratt has recently made this dimer in the gas phase, and determined the geometries of the two hydrogen bonds from the analysis of its electronic spectrum.

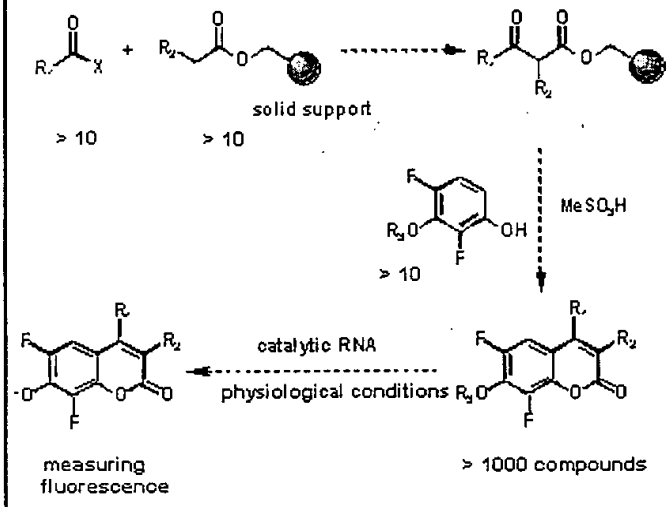
Hydrogen bonds (HB's) play a key role in many molecular recognition and assembly processes. Most HB's are asymmetric, with the hydrogen atom closer to the base, rather than the acid. But symmetric HB's are known, and believed to be important in enzyme catalysis (1). This project will employ the recently developed technique of high resolution electronic spectroscopy in the gas phase (2) to measure the moments of inertia of 1-benzoylacetone (1BA) and its deuterated analogs (3). Analyses of these data will yield the center-of-mass coordinates of the hydrogen atom to an accuracy of +/- 0.02 A, thereby showing whether or not the HB in 1BA is symmetric. Time permitting, applications to other systems might also be proposed. [2/02]

1. W.W. Cleland and M.M. Kreevoy, Science 264, 1887 (1994).
2. W.A. Majewski, et al., Laser Techniques in Chemistry 23, 101 (1995).
3. B. Schiott, et al., J. Am. Chem. Soc. 120, 12117 (1998).

## Professor Kazunori Koide

### *A Combinatorial Approach Towards The Development of a High-throughput Assay System to Measure RNA Levels In Vivo*

To develop a sensitive method to quantify primary RNA transcripts *in vivo*, a system will be developed in which RNA itself catalytically generates fluorescent molecules so that the fluorescent signal is amplified. A combinatorial library of coumarin derivatives will be prepared on solid support according to scheme 1. After the library of DNA is transfected into yeast, each coumarin derivative will be incubated in an individual well combined with the yeast library. If an RNA molecule catalyzes the cleavage of ArO-$R_3$ bond to generate fluorescence, the yeast cells will be isolated by means of a fluorescence-activated cell sorter.

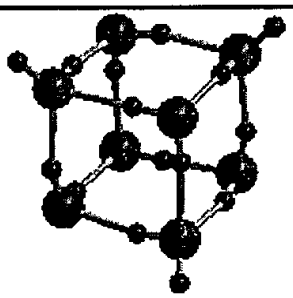



Our approach is flexible in that we will be able to manipulate both fluorescent molecules and RNA by means of synthetic organic chemistry and molecular biology. This flexibility would increase the feasibility of the proposed project exponentially. This proposed method would be facile and allow for high-throughput screenings to identify novel small molecules that regulate gene expression at the transcription level. Since the readout does not involve other steps besides transcription in gene expression, one will be able to observe transcriptional activities without disrupting living cells.

A REU student is expected to carry out synthetic studies for the preparation of combinatorial library compounds.[1/02]

## Professor Kenneth Jordan

### *Computer "Measurements" of the Properties of Atomic and Molecular Clusters*



Computed structure of the cubic form of the water octamer.

Computer simulations permit "measurements" on complex chemical systems that would be difficult or expensive to characterize experimentally.
In this project the student will use Monte Carlo or molecular dynamics computer simulation methods to characterize the melting behavior of neutral and charged water clusters. [1/02]

## Professor Kenneth Jordan

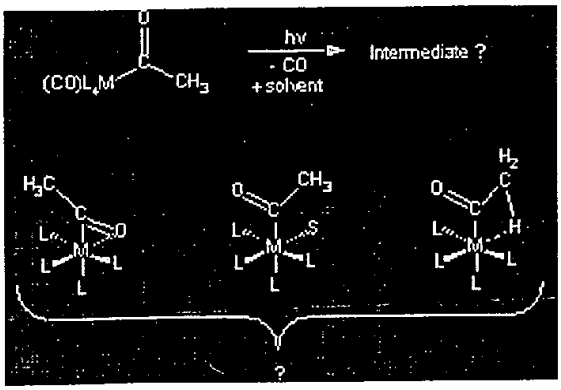
### *Computer measurements of Protein Folding*

Computer simulations have proven very valuable in providing insight into the protein folding process. In this project the student will explore the use of new optimization methods for locating the global minimum structure of models for protein folding. The project is flexible in the sense that the student can either use computer codes already developed within our group or can become involved in extending the codes to be more flexible. [1/02]

## Professor Joseph Grabowski

### *Measuring Activation Volumes for Reactive Intermediates: Application to Mechanism Determination in C-H Bond Activation*

Recently, a new tool, Photoacoustic Calorimetry (PAC) has emerged as a reliable methodology which allows for the characterization of intermediates in at least some reactions.(Ref.




The activation of C-H bonds remains a potentially useful target in the area of organometallic chemistry. The design of catalysts, such as those necessary for C-H bond activation can be approached in at least two different ways, either by empirically trying compounds or by elucidating the detailed mechanism and from it, selecting the appropriate catalytic molecule to prepare and use. A major limitation in mechanism elucidation are tools for the characterization of the key reactive intermediates along the reaction coordinate.

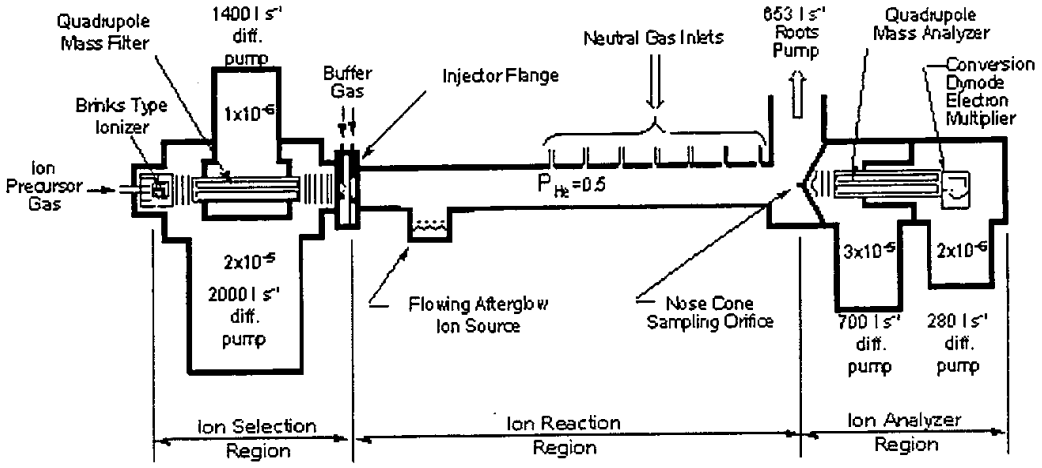
1) In this project, the student will use the custom-built PACs in the Grabowski group to measure the volume changes associated with photodissociation of CO from selected metal carbonyls (e.g., $CpMn(CO)_3$, $Cr(CO)_6$, etc.) in order to create a data base of values that can be used to interpret the meaning of a similar measured volume change for a proposed intermediate in an actual C-H activation catalyst. The student will conduct a series of PAC experiments in an homologous series of solvents, will experimentally measure the absolute quantum yield in one of those solvents and relative values in the remaining solvents, and then will interpret their observations, using custom-designed data alogrithims, in terms of proposed intermediates. Simultaneous to making a series of much-needed measurements, the student will also have the opportunity to contribute to the refinement of this newly emerging, general purpose measurement technique, either through instrumental or data reduction modifications brought about by an understanding of their problem and the operation of the existing instruments and programs.(Ref. 2) [1/02]

1. Listening to Reactive Intermediates: The Application of Photoacoustic Calorimetry to $B_{12}$", R. R. Hung; J. J. Grabowski *J. Am. Chem. Soc.*, 1999, *121*, 1359-1364
2. "Photoacoustic Calorimetry: An Advanced Undergraduate Physical-Organic Chemistry Laboratory", B. Fletcher; J. J. Grabowski *J. Chem. Ed.*, 2000, 77, 640-645.

## Professor Joseph Grabowski

### *Measuring the Specificity and Rapidity of Ion-Molecule Reactions for Real-Time Complex Mixture Analysis.*

Gas-phase ion chemistry has contributed enormously to our basic understanding of fundamental organic chemistry properties and reactions, especially since the landmark papers by Brauman on the acidities of the alcohols and basicities of the amines.(Ref 1) We are now well positioned to use the chemical understanding that has accumulated, along with the unique instrumental capabilities of the Flowing Afterglow to develop applied, analytical applications for the real-time, quantitative analysis of complex mixtures of volatile organic compounds. For example, patients diagnosed with schizophrenia seem to have substantially higher amount of carbon disulfide on their breath than others, suggesting that it is being synthesized somehow. However, these early studies were carried with cumbersome chemical procedures necessarily limiting the amount of data that was collected. Can we design specific chemical reactions, which combined with the unique capabilities of the Flowing Afterglow, can be used to identify and quantitate carbon disulfide, in real time, in a complex mixture, without requiring chromatography and without requiring calibration curves? The student working on this project will use our laboratory based instrument (the SIFT, Figure; Ref. 2) to measure the specificity and rapidity of some novel ion-molecule reactions in order to evaluate their potential usefulness in for carbon disulfide detection. Once ideal reactions are identified and characterized, they will then be applied to complex mixtures to establish reliability and detection limits for authentic samples. [1/02]

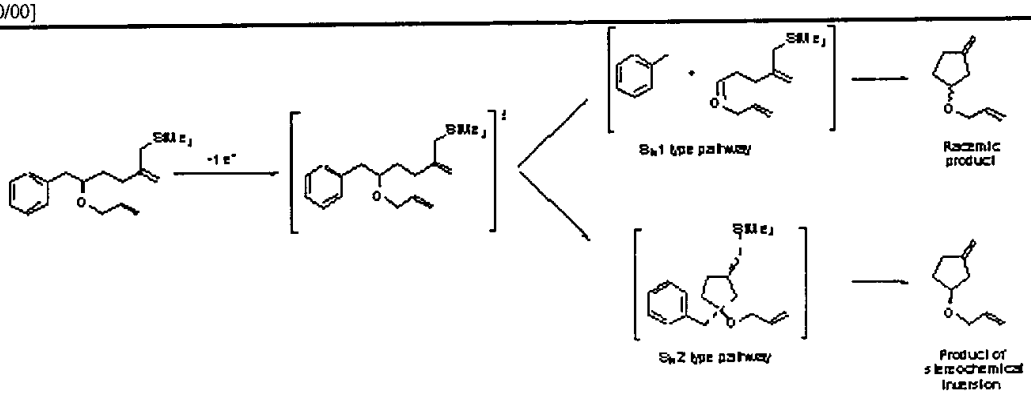



1. Brauman, J. I.; Blair, L. K. *J. Am. Chem. Soc.* **1969**, *91*, 2126-2127. Brauman, J. I.; Riveros, J. M.; Blair, L. K. *J. Am. Chem. Soc.* **1971**, *93*, 3914-3916.
2. "Simplified Injector Flanges for the Selected Ion Flow Tube", V.N. Fishman and J.J. Grabowski, *Int. J. Mass Spectrom.*, **1998**, *177*, 175-186. "Selected Ion Flow Tube Studies of the Atomic Oxygen Radical Cation Reactions with Ethylene and Other Alkenes", V.N. Fishman, S.T. Graul, and J.J. Grabowski, *Int. J. Mass Spectrom.*, **1999**, *185/186/187*, 477-496.

## Professor Paul Floreancig

### *Measuring Stereochemistry to Aid Mechanistic Studies of the Single Electron Oxidation Initiated Cyclization*

In this project the student will make measurements designed to explore the mechanism of a new chemical reaction, the single electron oxidation initiated cyclization, by determining the extent to which stereogenicity can be retained when an enantiomerically enriched substrate undergoes cyclization. Two mechanistic extremes can be envisioned for this transformation (shown below). The student working on this project will be expected to prepare enantiomerically enriched substrates and then subject them to the cyclization conditions. The extent to which each mechanism contributes to the overall process can be determined by comparing the enantiomeric excesses of the products to those of the substrates versus that predicted for each mechanism. Among other things, this project will provide the student with experience in the use of spectroscopic and/or chromatographic techniques for the measurement of enantiomeric excesses, and in the design of simple experiments directed toward the development of a mechanistic model for an organic reaction. [10/00]




## Professor Dennis Curran

### *Synthesis and Measurements of New Fluorous Reagents*

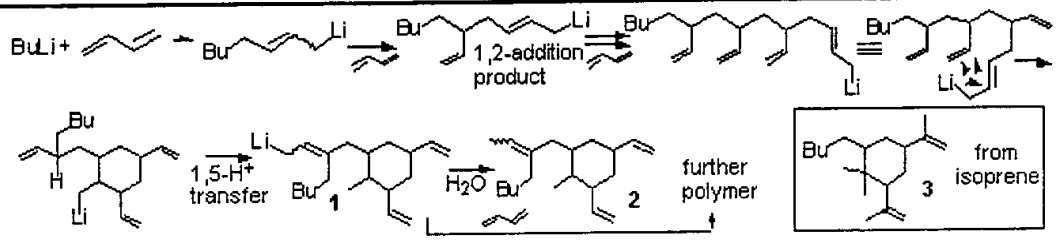
Recently developed techniques for "fluorous synthesis" unite the reaction and separation processes in organic reactions (Refs. 1,2). Molecules bearing fluorous (highly fluorinated) tags can be separated from organic (non-tagged) molecules by fluorous liquid-liquid or solid liquid extraction. However, because the field is very young, there are relatively few fluorous reagents available to pair with the new separation techniques. This project will involve the synthesis of a new fluorous reagent or protecting group and the study of its use in representative organic transformations. [1/02]

1. A. Studer; S. Hadida; R. Ferritto; S. Y. Kim; P. Jeger; P. Wipf; D. P.Curran, *Science* **1997**, *275*, 823-826.
2. Luo, Z. Y.; Zhang, Q. S.; Oderaotoshi, Y.; Curran, D. P. "Fluorous mixture synthesis: A fluorous-tagging strategy for the synthesis and separation of mixtures of organic compounds" *Science* **2001**, *291*, 1766-1769.

## Professor Ted Cohen

### *Using Spectroscopic Measurements to Determine the Products of Alkyl Lithium Induced Diene Polymerizations*

In a recent study of the lithium ene-cyclization, the intramolecular addition of an allyllithium to an unactivated alkene, we have learned that the reaction is far more facile than the better known magnesium ene-reaction but, unlike the latter, it is thermodynamically unfavorable and will only be observed when it is followed by an irreversible consummating reaction such as a 1,5-transfer of an allylic proton. This new insight allows for the first rational explanation of the fact that polymerization of butadiene or isoprene, initiated by butyllithium in the presence of tetramethylethylenediamine (TMEDA) under conditions that lead mainly to 1,2-addition, produces polymer with a substantial percentage of ring structure provided that the butadiene is fed in slowly. A number of very unsatisfactory explanations have been put forth in papers and patents. We surmise that in the case of butadiene the process shown in the scheme is occurring to form 1. In the case of isoprene, the 1,5-proton transfer is from an allylic methyl and an important intermediate would be protonated in workup to produce 3. The undergraduate research project would consist of performing oligomerizations using 3 moles of diene per mole of butyllithium. In the absence of excess diene, 2 or 3 would be produced. Various measurement techniques would be used to prove their structures. These will include 1D and 2D NMR spectra, IR spectra, mass spectra, unsaturation determination, and x-ray crystal structural measurements of a derivative. [10/00]




## Professor Rob Coalson

### *Computer-based Measurement of Electric Current through Nanostructures*

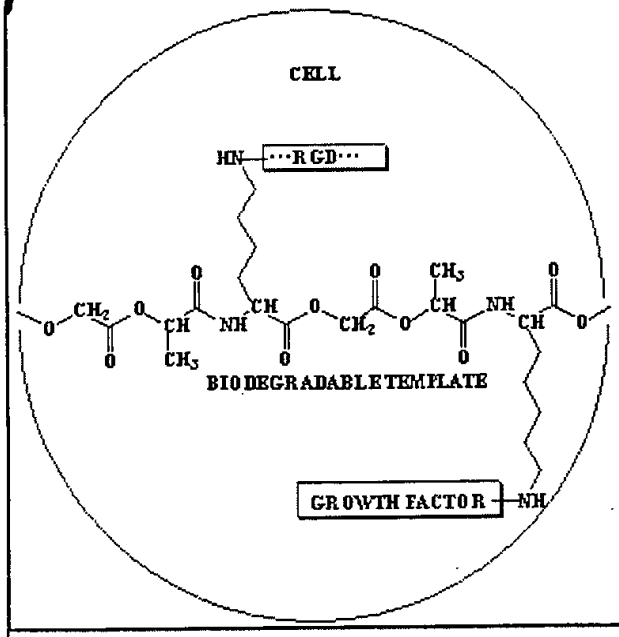
The Coalson group performs calculations of charge transfer through both biological and synthetic nanostructures. In paricular, we study the transport of ions like Na+ and Cl- through protein channels which are embedded in cell walls.

We also study electron transport through organic molecular wires. Fundamental concepts from theoretical chemistry are converted into computer programs which are run on large-scale computer platforms such as Pitt's Center for Molecular and Materials Simulation. These calculations aim to increase understanding of the charge transfer processes involved, and ultimately to enable improved design of both biological and synthetic nanowires and nanopores. Undergraduates with a strong interest in computing and physics, as well as chemistry, may find this line of research appealing.

## Professor Toby Chapman

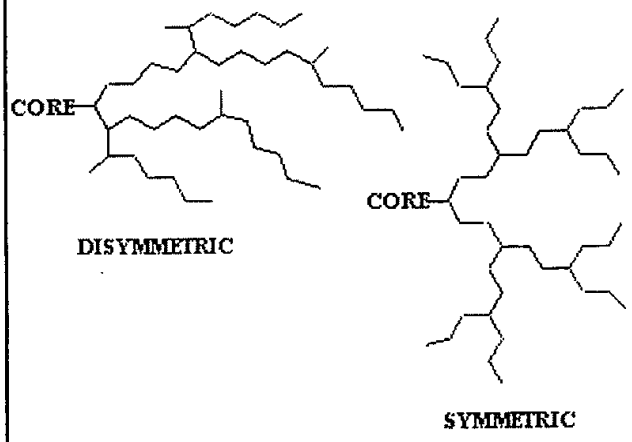
### *Measuring the Effects of Covalently Attached Growth Factors in New Tissue Engineering Templates*

We are interested in making biodegradable polymer templates for tissue engineering that contain functional groups allowing for attachment of recognition peptides and growth factors. We propose to make initiator molecules containing protected amines from which can be grown polylactic-polyglycolic acid copolymers or polycaprolactone. The amines would be unmasked under mild conditions and some short peptides would be attached. We would then test these for their competence in growing osteoblast precursor cells. We would also test the effect of covalently attaching growth factors know to be important in cell growth. [3/02]




## Professor Toby Chapman

### *Measuring Solution Property and Surfactant Ability Differences in Symmetric and Asymmetric Isomeric Dendrimers.*

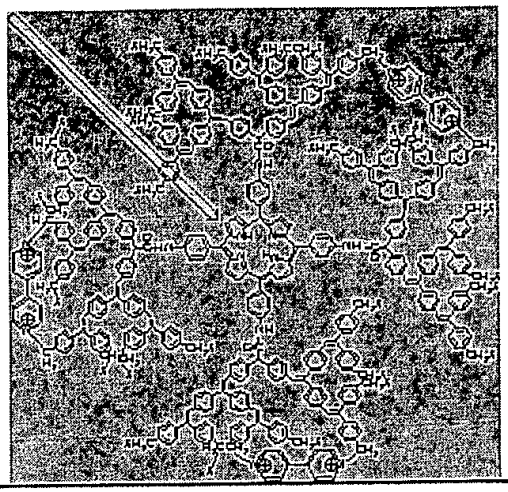



Dendrimers are highly branched, uniform polymers with a number of very interesting properties. We have been studying amphiphilic poly-L-lysine dendrimers for their surfactant and emulsifier properties with their possible use in controlled drug delivery. Most of the dendrimers currently studied are symmetrical molecules with symmetry at each branching point. Polylysine, with both α-, and ε-amines is not symmetric. Symmetry is believed to be important in some of the special solution properties of dendrimers but this, in fact, remains unstudied. We wish to synthesize a symmetric, achiral isomer if lysined, 4-amino-2-(2-aminoethyl)butanoic acid and prepare the corresponding dendrimers. These would be tested for their solution properties and surfactant properties to the lysine dendrimers and in this way test the influence of dendrimer symmetry, if any, on these properties. We will compare the ability of these molecules to solubilize hydrophobic medications as well. [3/02]

## Professor Toby Chapman

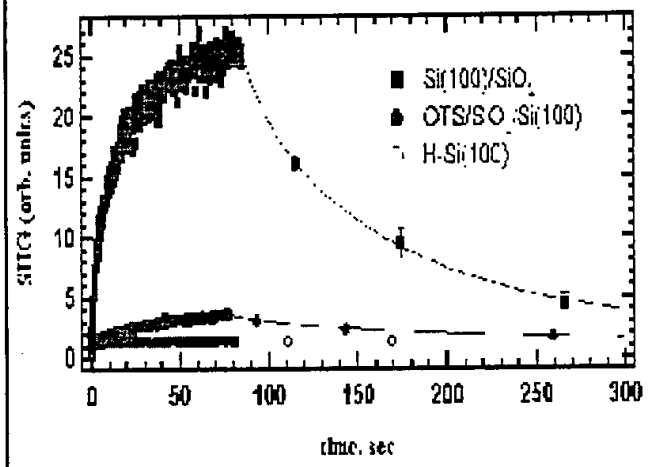
### *Measuring the Light Emission and Non-linear Optical Properties of a New Conjugated Dendritic Polymer.*

Conjugated dendrimers are rare, only one reported in the literature. A conjugated poly(phenylene vinylene) dendrimer would be of interest to compare with the known linear polymer. The latter emits light under the influence of an electric field and is a conductor of electricity after "doping". The comparision with the dendritic isomer would be of interest. Also, these compounds should show the property of harvesting light and concentrating the energy at the central core. We propose to use these properties to make a photoredox catalyst as well as to create an organic quantum dot with strong non-linear optical properties.. [3/02]



## Professor Eric Borguet

### *Measuring and controlling charge transfer and trapping processes at semiconductor interfaces with nanotechnological applications*




Charge transfer dynamics at a semiconductor interface. Transient SHG response depends on surface chemical termination.

With the ultimate goal of integrating molecular electronics with semiconductor electronics for nanotechnological applications, undergraduate researchers, working with advanced graduate students, will construct chemically modified semiconductor interfaces and probe charge transfer and trapping process at these interfaces. Our strategy is to design organic monolayers decorated with chemical functional groups that can provide either electron donating or accepting capability. Nonlinear optical techniques, that provide surface sensitivity are used to probe the charge transfer processes that may be initiated by a variety of means including optical, chemical and electrochemical perturbation. Students will have the opportunity to explore synthesis as well as state-of-the-art ultrafast laser spectroscopy. A variety of other probes including Atomic Force Microscopy and Scanning Tunneling Microscopy are available to characterize the surfaces on the nanometer scale. [2/01]

## Professor Sanford Asher

### *Direct Spectroscopic Measurements of the First Steps in Protein Folding*

With the completion of the determination of the human genome we will know the primary structure of the ~100 K proteins we make in order to live. Most human disease results from errors in the synthesis of proteins. It is now possible to determine the primary structure differences between the diseased protein and the normal protein. Strategies for disease treatment could be developed if the structure and function of the diseased and normal proteins were known. Unfortunately, the structure and function of over 95,000 of these proteins are unknown. Although it should be possible to calculate the structure and function of proteins from their primary sequence, the laws of protein folding are as yet unknown. My group is determining the rules of protein folding using measurements made with powerful laser techniques. This research program offers you the opportunity to have a positive impact on the human condition while learning both spectroscopy and biology. [1/02]

## Professor Sanford Asher

### *Measurements to Optimize Intelligent Materials for Chemical Sensing and Optical Switching*

My group has developed novel smart materials whose optical properties can be altered by chemical analytes and light. These materials are being developed for use as optical switching transistors for the next generation of optical computers. In addition, these materials are being developed for use as *in vivo* glucose sensors for diabetic patients. The idea is to implant this sensor under the skin and to monitor its color in order to detect the glucose concentration.

This is a research program which will teach students how to make measurements on prototype materials and then use the results of these measurements in conjunction with aspects of polymer synthesis, materials science and optics to refine the sensor under development. [1/02]

## Professor Shigeru Amemiya

***Measuring Surface pH by Electrochemical Nanosensors***

pH is an important parameter that determines the extent and rate of many chemical and biological reactions. Glass electrodes, therefore, have been widely used for decades to measure pH in bulk solutions. However, when such reactions occur at a surface, pH changes only near the surface. Surface pH, for instance, can be changed by transport of $H^+$ through ion channels across biomembranes and by corrosion at metal surfaces. In order to quantitatively study surface reactions of biological and industrial importance, we will fabricate and characterize electrochemical pH sensors in submicrometer and nanometer dimensions. The pH sensors will be fabricated by using different materials such as polymer membranes and metal oxides. The sensors will be used as a probe of scanning electrochemical microscopy to measure surface pH in situ. For more information, please visit my web page by clicking my name!. [2/02]

REU Web page maintained by Prof. Joseph J. Grabowski. Updated 03/05/2002 17:12:10

Fluka & Riedel Home

Site Map

About Fluka and Riedel

+ Fluka and Riedel Quality

− New Products

2003

+ Analytical

+ Bioscience

+ Organic / Synthetic

+ Specialties

+ Fine Chemicals / Bulk

CustomWare / OEM

+ Literature

+ Miscellaneous

Contact Us

Feedback

# New Products Synthetic Chemistry First Quarter 2003

## New Nitrilases for Selective Nitrile Hydrolysis

Fluka offers several nitrilases for enzyme catalyzed reactions. Biocatalytic hydrolysis of nitriles to carboxylic acids can be carried out under mild conditions, without affecting other functional groups in high regio- and stereoselectivity. The nitrile group thus functions as a "masked" acid group.

**Enzymes for Organic Synthesis are a part of Fluka`s approach to "Green Chemistry".**

**Table 1. New Nitrilases for Selective Nitrile Hydrolysis**

| Catalog No. | Product Name | Activity | Package Size |
|---|---|---|---|
| 53841 | Nitrilase, Arabidopsis thaliana, recombinant from E. coli[1] | ~ 0.5 U/mg[A] | 10 mg, 50 mg |
| 82429 | Nitrilase from Alcaligenes faecalis[2] | ~15 U/g[A] | 10 mg, 50 mg |
| 78424 | Nitrilase from Pseudomonas fluorescens[3] | 10–15 U/g[B] | 10 mg, 50 mg |
| 76713 | Nitrilase from Rhodococcus rhodochrous[4] | ~10 U/g[C] | 10 mg, 50 mg |
| 72295 | Nitrilase from Rhodococcus sp. | >0.1 U/g[C] | 10 mg, 50 mg |

- [A] 1U corresponds to the amount of enzyme which liberates 1 µmol ammonia per minute at pH 8.0 and 35°C with the conversion of 3-phenylpropionitril to 3-phenylpropionic acid
- [B] 1U corresponds to the amount of enzyme which liberates 1 µmol ammonia per minute at pH 7.5 and 30°C with the conversion of 2-thiophenacetonitrile to 2-thiopheneacetic acid
- [C] 1U corresponds to the amount of enzyme which liberates 1 µmol ammonia per minute at pH 8.0 and 30°C with the conversion of benzonitrile to benzoic acid

### References

[1] Catalyst for the (E)-selective hydrolysis of (E,Z)-α,β-unsaturated nitriles to carboxylic acids: F. Effenberger, S. Osswald, *Tetrahedron: Asymmetry*, 12, 2581 (2001); Selective hydrolysis of aliph. dinitriles to monocarboxylic acids: F. Effenberger, S. Osswald, Synthesis, 1866 (2001); enantioselective hydrolysis of (1)-arylacetonitriles: F. Effenberger, S. Osswald, *Tetrahedron: Asymmetry*, 12, 279 (2001).

[2] A novel nitrilase: T. Nagasawa *et al.*, *Eur. J. Biochem.*, 194, 765 (1990).

[3] Enantioselective hydrolysis of O-acetylmandelonitrile: N. Layh *et al.*, *Arch Microbiol.*,158, 405 (1992).

[4] Enantioselective synthesis of acids by hydrolysis of nitriles: M. L. Gradley-L, C. J. Knowles, *Biotechnol Lett*, 16, 41 (1994); T. C. Bhalla *et al.*, *Appl. Microbiol Biotechnol.*, 37, 184 (1992).

Back to Top

Back to New Products home

# 2 Synthetic methods Part (iv) Protecting groups

**Alan C. Spivey and Benjamin I. Andrews**

*Department of Chemistry, Brook Hill, University of Sheffield, Sheffield, UK S3 7HF*

Another excellent and comprehensive "update" review of protecting group (PG) strategies in organic synthesis has appeared this year.[1] Additionally, a substantially expanded and updated 3rd edition of the classic "*Protective Groups in Organic Synthesis*" has been published.[2] An introductory text on protecting groups has also appeared.[3]

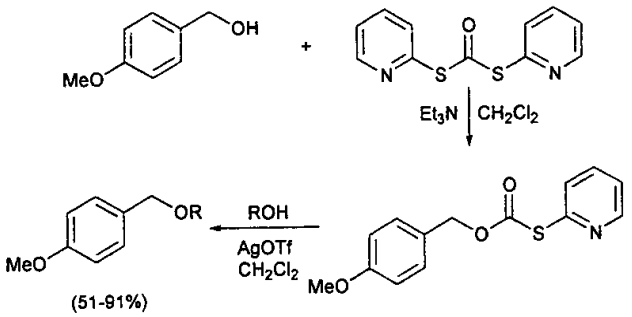
## 1 Hydroxy protecting groups

Simple alkyl ethers provide limited opportunities as PGs due to the harsh conditions generally required for their deprotection. However, an efficient CsOH–tetra-*n*-butylammonium iodide (TBAI)–4 Å molecular sieves (MS) variant of the Williamson etherification should find widespread utility for the introduction of ether based PGs.[4] The cyclohexyl (Chx) ether has been advocated as a particularly robust group for serine side-chain protection during *tert*-butoxycarbonyl (Boc) based Solid Phase Peptide Synthesis (SPPS).[5] Introduction is *via* reaction of the sodium alkoxide of *N*-Boc serine with 3-bromocyclohexene then hydrogenation with Adams' catalyst ($PtO_2$). Removal is by treatment with 1 M trifluoromethanesulfonic acid (TfOH)–thioanisole in trifluoroacetic acid (TFA). The use of TBAI in combination with $BCl_3$ in $CH_2Cl_2$ has been shown to dramatically enhance the rate of cleavage of primary alkyl aryl ethers to give phenols (Scheme 1).[6] The reaction takes 1–2 h at −78 °C and secondary alkyl ethers are untouched.

Demethylation of asymmetrically substituted 2,6-dimethoxybenzaldehydes with $MgI_2{\cdot}Et_2O$ has been shown to proceed regioselectively at the more sterically hindered methoxy group.[7]

Incremental improvements, particularly with respect to environmental impact, to procedures for the introduction and deprotection of acetal based PGs such as tetrahydropyranyl (THP) ethers continue to be reported. Of note for introduction are the use of $H_2SO_4$ adsorbed on silica ($SiO_2$) with microwave (MW) irradiation in the absence of solvent,[8] and the use of $SnCl_2{\cdot}2H_2O$ in $CHCl_3$.[9] Montmorillonite

DOI: 10.1039/b003402i 




**Scheme 2**

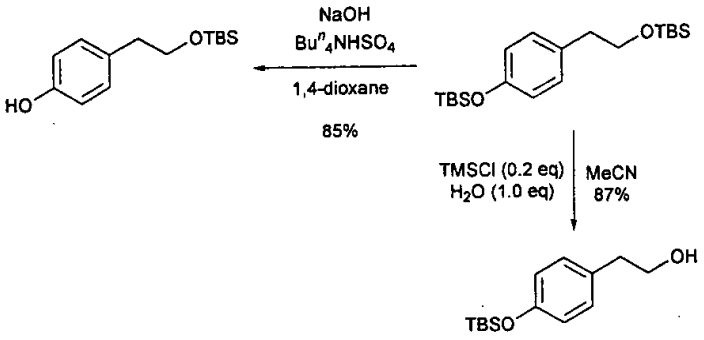
amino)pyridine (DMAP), 0 °C to rt] TBSCl (1 equiv.) and TrCl (1 equiv.) give opposite selectivities: TBS ether formation occurs at the primary alcohol, trityl ether formation occurs at the phenol.[24] The trityl selectivity is preserved irrespective of the nature of the alkyl alcohol but the TBS selectivity is eroded for secondary alkyl alcohol *vs.* phenol competition. Two new methods have been described for the introduction of the *p*-methoxybenzyl (PMB) group under essentially neutral conditions. Primary and secondary alcohols can be protected using PMB-OH, $Yb(OTf)_3$ (5 mol%) in $CH_2Cl_2$ at rt in the presence of THP and TBS ethers, benzoyl (Bz) esters, and acetonides.[25] A more powerful method that also allows protection of tertiary alcohols uses 4-methoxybenzyl 2-pyridylthiocarbonate (PMB-TOPCAT) in $CH_2Cl_2$ with stoichiometric AgOTf as promotor (Scheme 2).[26]

A new method for cleavage of benzyl ethers (and benzylidene acetals) uses $NaBrO_3$–$Na_2S_2O_4$ in EtOAc–$H_2O$.[27] This reagent combination effects rapid radical benzylic bromination at rt followed by hydrolysis of the resulting α-bromo ether. Acetyl, chloroacetyl, benzoyl, pivaloyl, tosyl, TBS, trityl, and acetonide groups remain unaffected. For removal of PMB ethers, methods involving 80% aq. AcOH at 80 °C,[28] and $CeCl_3 \cdot 7H_2O$–NaI in MeCN at reflux[29] have appeared. Both systems are selective in the presence of methyl ethers, benzyl ethers, benzyl esters, and acetates. The latter system also displays tolerance of alkenes but cleaves allyl ethers.[30] For removal of *p*-nitrobenzyl (PNB) ethers (and esters) the use of indium in EtOH (aq.)–$NH_4Cl$ offers a simple alternative to $Na_2S$ or $Na_2S_2O_4$ based reductive methods.[31] Three new benzylic PGs have been introduced: the *p*-(dodecyloxy)benzyl group to facilitate purification of oligosaccharides on $C_{18}$ silica,[32] the *p*-acetoxybenzyl (PAB) group, and the 2-(trimethylsilyl)ethoxymethoxybenzyl (*p*-SEM-benzyl) group.[33] Both these latter groups are orthogonal to simple benzyl and PMB ethers. Deprotection is effected by removal of the "phenolic" PG [NaOMe, and tetrabutylammonium fluoride (TBAF), respectively] followed by heating to 65 °C or oxidation using *e.g.* 2,3-dichloro-5,6-dicyanobenzoquinone (DDQ), $FeCl_3$, $PhI(OAc)_2$ or $Ag_2CO_3$ on $SiO_2$. Deprotection of allyl ethers (and esters) has been described using TMSCl–NaI.[34] The method is compatible with acetonide protection but is unlikely to be selective relative to many other PGs including alkyl esters and silyl ethers. An interesting method for "one-pot" deprotection of aryl allyl ethers employs an electrochemically gener-

sensitive functionality. Iminophosphorane bases [*e.g.* $BnN{=}P(MeNCH_2CH_2)_3N$], and a polymer bound variant, are also efficient catalysts for acylation of primary, but not secondary, alcohols with enol esters in THF at rt.[50] Notably, acid labile groups such as acetals and epoxides, and groups such as TBS ethers and disulfides, which are cleaved by *e.g.* $Sc(OTf)_2$–$Ac_2O$, are unaffected. $In(OTf)_3$ has been advocated as an inexpensive alternative to $Sc(OTf)_2$ for acylation of primary and secondary alcohols using $Ac_2O$.[51] A review detailing strategies for the selective esterification of the three secondary hydroxy groups present in cholic and related bile acids highlights some interesting nuances in reactivity towards acylation as a function of the steric and stereoelectronic environment of the relevant hydroxy groups.[52] β-Oxypropyl formate, which is prepared from propargyl alcohol and formic acid (1:1) in the presence of $RuCl_2(PPh_3)$(*p*-cymene) (1 mol%), has been introduced as a mild formylating reagent for secondary alcohols.[53] The reaction is carried out at 70 °C in THF with 1,5-diazabicyclo[4.3.0]non-5-ene (DBN, 10 mol%) as catalyst. The direct conversion of primary and secondary triethylsilyl (TES) and TBS ethers into formate esters using the Vilsmeier–Haack complex [$POCl_3$–DMF] in DMF at rt has also been described.[54] $Yb(OTf)_3$ has been found to catalyse deacetylation of phenols and primary alcohols in the presence of secondary acetates, benzoates, *p*-nitrobenzoates, and pivaloates.[55] A number of reports have detailed selective introduction and removal of ester PGs in complex carbohydrates and nucleosides using enzymes.[56–58]

The trichloroacetimidate group has been widely employed as an excellent leaving group for the preparation of cations (*e.g.* benzyl and *tert*-butyl, for esterification purposes) and as a neighbouring group nucleophilic nitrogen in Overman type rearrangements. However, it can be employed as an alcohol PG which is orthogonal to acetates and TBS ethers.[59] Introduction involves trichloroacetonitrile–1,8-diazabicyclo[5.4.0]undec-7-ene (DBU). Deprotection in the presence of acetates involves toluene-*p*-sulfonic acid monohydrate ($TsOH{\cdot}H_2O$, 0.5 equiv.) in $CH_2Cl_2$–MeOH (1:1) at rt for 5 min or Zn dust–$NH_4Cl$ in EtOH at reflux for 5 min. Deprotection in the presence of TBS ethers is achieved using DBU (1 equiv.) in MeOH. The trichloroacetimidates are stable to TBS removal with tetrabutylammonium fluoride (TBAF).

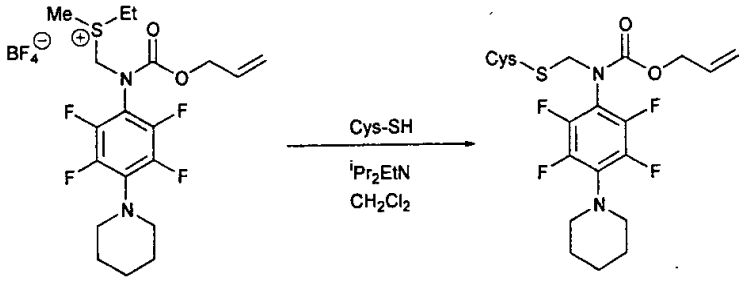
New conditions for sulfonate ester protection/activation of alcohols with sulfonyl chlorides employing $Me_3N{\cdot}HCl$ (10 mol%) as a "co-catalyst" with $Et_3N$ have been described.[60] The co-catalyst can either be employed with 1 equiv. of $Et_3N$ or with 10 mol% of $Et_3N$ and using KOH or $Ca(OH)_2$ (1 equiv.) as stoichiometric base in $CH_2Cl_2$ or toluene. It is proposed that the $Et_3N$ rapidly deprotonates the $Me_3N{\cdot}HCl$ such that $Me_3N$ can act as an efficient nucleophilic catalyst. The reductive cleavage of sulfonate esters by $LiAlH_4$ can occur by S–O bond cleavage to give the corresponding alcohol and a sulfinic acid or by C–O bond cleavage to give the deoxygenated substrate and a sulfonic acid. Evidence has been presented that in the case of mesylates the former mode (which predominates for secondary and tertiary mesylates) occurs by elimination of a sulfene (E1cB) whereas the latter mode (which predominates for primary mesylates) occurs by direct substitution ($S_N2$).[61] S–O Bond cleavage of sulfonate esters (and sulfonamides) can also be achieved with MW irradiation using 37% KF impregnated $Al_2O_3$.[62] The method




**Scheme 4**

A neat method for alkyl carbonate protection of alcohols, which avoids the use of phosgene or its derivatives, is the $Cs_2CO_3$ promoted *O*-alkylation of alcohols using the appropriate alkyl bromide and TBAI in DMF at 90 °C.[81] The method could find utility in *e.g.* the preparation of a new photolabile carbonate PG: the 2-(2-nitrophenyl)ethoxycarbonyl group. This has been piloted for 5′-protection during oligonucleoside synthesis and is deprotected by irradiation at 365 nm in MeOH–$H_2O$.[82] However, progress has been made in defining efficient electron transfer sensitizers for photolytic deprotection of the parent phenacyl carbonates for the same purpose.[83] Thus 9,10-dimethylanthracene and 9-methylcarbazole sensitizers allow efficient deprotection of 5′-phenacyl carbonates with irradiation at 398 and 345 nm respectively. Selective protection of the hydroxy group of α-hydroxy acids as an ethoxycarbonyl group can be efficiently accomplished using 2-ethoxy-1-(ethoxycarbonyl)-1,2-dihydroquinoline (EEDQ) in $CH_2Cl_2$ at rt.[84] Deprotection is readily accomplished using *e.g.* 1% $K_2CO_3$ in MeOH.

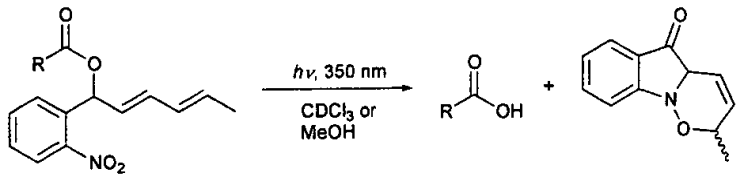
For the deprotection of acetonide protected 1,2-diols an interesting procedure employing thiourea in EtOH–$H_2O$ has been described.[85] It is noteworthy because of the essentially neutral conditions of the transformation, in contrast to most alternatives which employ acids. The conditions are also appropriate for unmasking of 1,3-dioxolane protected ketones and aldehydes. Acetonide protected 1,2-diols having a neighbouring heteroatom (*e.g.* hydroxy, alkoxy, amino, but not thiol) can be partially deprotected to the corresponding 1,2-hydroxy-*tert*-butyl ethers using MeMgI.[86] Chelation plays a key role in directing the regioselectivity of ring-opening such that the unmasked hydroxy group is adjacent to the chelating group (Scheme 5).




**Scheme 6**

## 3 Carboxy protecting groups

Esters remain the most popular method for protection of carboxy functionality during synthesis. For their preparation from an appropriate alcohol, a new dehydrating reagent, 2-chloro-1,3-dimethylimidazolinium chloride (DMC), appears to be an attractive alternative to traditional coupling reagents such as *N*,*N*′-dicyclohexylcarbodiimide (DCC).[91,92] The reagent is a mildly moisture sensitive crystalline solid which is prepared from commercial 1,3-dimethylimidazolidin-2-one (DMI) by chlorination with oxalyl chloride. Typical conditions for esterification use equimolar amounts of both acid and alcohol, DMC (1 equiv.), pyridine (2 equiv.) in $CH_2Cl_2$ at rt. Even highly hindered esters (*e.g. tert*-Bu) esters are readily prepared by this method. For the synthesis of methyl esters by *O*-alkyl bond formation a new method using $LiOH{\cdot}H_2O$–$Me_2SO_4$ offers a cheap and effective alternative to the popular diazomethane based protocol.[93] The carboxylic acid is reacted with the $LiOH{\cdot}H_2O$ in dry THF and the resultant carboxylate reacted with $Me_2SO_4$ (0.5–1 equiv.) at reflux for 0.5–3 h. Excellent chemoselectivity was observed in the presence of phenols, amines, amides, and *N*-Boc protected amines. *N*-Boc α-amino acids could be protected at rt and without racemisation. A number of new ester groups offering advantages for specific applications have been reported this year. The heptafluorotriphenylmethyl ($TrtF_7$) ester has been developed for side-chain carboxy-protection of aspartic and glutamic acids during Fmoc based SPPS.[94] This has been designed to complement trityl ether side-chain protection of serine, threonine and cysteine and allow for side-chain deprotection under mildly acidic conditions (1% TFA in $CH_2Cl_2$). The group is introduced using heptafluorotrityl chloride–DIPEA. The heptafluorotrityl chloride is prepared by chlorination of heptafluorotrityl alcohol which in turn is prepared from 4,4′-difluorobenzophenone and pentafluorophenylmagnesium bromide. Fluoren-9-ylmethyl (Fm) esters are occasionally employed for Boc based SPPS in the N to C direction. Existing methods for their preparation employ DCC–DMAP coupling of the acid, or imidazole mediated transesterification of *p*-nitrophenyl esters with fluoren-9-ylmethanol, or reaction of the acid with diazofluorene. A new method involves the use of




**Scheme 7**

hetero-Diels–Alder reaction with the nitroso group thereby "disarming" the side product (Scheme 7). The concept appears to work well but the pendent allylic/benzylic pentadienyl side chain looks rather reactive in the context of complex synthesis and will certainly be very acid sensitive.

An orthogonal set of allyl [Pd(0) labile], *tert*-Bu (TFA labile) and PNB (Zn–AcOH labile) esters has been used for the synthesis of 2-carboxysuccinic ester derivatives.[104] The orthogonal use of choline (enzyme labile) and allyl [Pd(0) labile] esters in the synthesis of base labile palmitoylated lipopeptides has been described.[105]

Due to the harsh conditions generally required for their hydrolysis, amides are not generally employed as carboxylic acid PGs. However, the novel use of the enzyme mushroom tyrosinase allows the use of acyl hydrazines as carboxylic acid PGs.[106] This commercially available enzyme, unlike many related oxidases, has good tolerance of organic solvents and effects oxidation of peptide derived acyl hydrazines to their corresponding acyl diazines in phosphate buffer–MeCN (or DMF) while bubbling oxygen at pH 7. The acyl diazines are hydrolysed *in situ* under these conditions. Certain classes of "amide" (*e.g.* Evans' oxazolidinones and Oppolzer's camphor sultam and related chiral auxiliaries) can be hydrolysed under conditions similar to those used for alkyl esters. It comes as no surprise therefore that related, achiral and chiral, thiazolidinethiones can be used as PGs for carboxylic acids and that they can be deprotected using DMAP promoted hydrolysis in MeCN at rt.[107,108]

## 4 Carbonyl protecting groups

Purification of aldehydes (and cyclic ketones) by crystallisation as their bisulfite adducts and then regeneration of the aldehyde, although well established, suffers from the harsh conditions required for the regeneration step (typically, rapid extraction into an organic phase from aqueous at pH ~10). New conditions which were developed to circumvent concomitant hydrolysis of a methyl ester under these conditions involve treatment of the bisulfite adduct with > 2 equiv. of TMSCl in anhydrous MeCN at ~50 °C and should greatly expand the scope of this purification/protection procedure.[109] Cyanohydrins and *O*-substituted cyanohydrins are frequently employed for protection of carbonyl groups from nucleophiles but generally display acute acid and Lewis acid instability. However,

## 5 Amine protecting groups

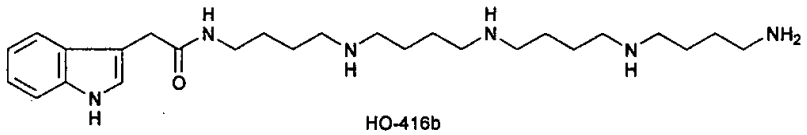
A new method for the direct conversion of alkyl and aryl azides into the corresponding acyl amines has been disclosed employing $Ac_2O$–TMSCl at reflux for ~1 h.[123] Alkyl esters and ethers remain intact during the process but phenols are acetylated. The procedure can also be used for trifluoroacetyl amide formation by using trifluoroacetic anhydride (TFAA). Selective mono-acylation of diamines is often problematic due to competing and often preferential di-acylation. A method for the selective benzoylation of primary amines in the presence of secondary ones involves initial treatment with 2 equiv. of BuLi, quenching with TMSCl (2 equiv.) to give the bis-TMS derivative, further deprotonation with BuLi (1 equiv.), and quenching with BzCl.[124] An aqueous work-up effects desilylation, affording the clean mono-benzoylated secondary amine.

The utility of dimethylmaleoyl protection of glucosamine as a means of promoting β-selective glycosylation (using the trichloroacetimidate–TMSOTf method) has been further demonstrated this year during the synthesis of lacto-*N*-neotetraose, a human milk constituent.[125] The group can be cleaved either using NaOH then dil. HCl or using hydrazine hydrate. A related PG for primary amines is the dithiasuccinoyl (Dts) group.[126] This group was developed for $N^{\alpha}$ protection of α-amino acids and is stable to strong acids and photolysis, but is rather base sensitive. It is introduced in a two-step process involving treatment with bis(ethoxythiocarbonyl) sulfide in EtOH–$H_2O$ (2:1) at pH 9–10 followed by (chlorocarbonyl)sulfenyl chloride (1.2 equiv.). This group has now been employed for primary amine protection during solid phase synthesis of peptide nucleic acid (PNA) oligomers wherein a detailed study delineated optimal conditions for deprotection as treatment with dithiothreitol (DTT, 0.5 M), in AcOH (0.5 M) in $CH_2Cl_2$ for 5–8 min. The group is compatible with *N*-[(1*H*-benzotriazol-1-yl)(dimethylamino)methylene]-*N*-methylmethaniminium hexafluorophosphate *N*-oxide (HBTU)–DIPEA (3:1) in NMP. Other protective groups for primary amines described this year are the *N*,*N*-dibenzyl- and *N*,*N*-dimethylformamidine groups[127] and the diphenylsilyldiethylene (DPSide)[128] group. *N*,*N*-Dimethylformamidines of *alkyl* amines are hydrolysed under neutral conditions (MeOH–$H_2O$, rt) to give the corresponding formamides exclusively whereas the corresponding *N*,*N*-dibenzylformamides cleanly afford the free amines under identical conditions. Hydrogenolysis of either group gives mixtures of *N*-methyl and free amines. However, formamidines find widespread use for the protection of *aromatic* amino groups, *e.g.* during nucleoside base manipulation, for which hydrolysis requires vigorous basic hydrolysis. Consequently, the finding that *N*,*N*-dibenzylformamidines of aromatic amines can be hydrogenated to give the free amines cleanly is significant. The DPSide group resists acidic, basic and hydrogenolytic conditions and is orthogonal to *N*-Boc, *N*-benzyloxycarbonyl (Cbz), and *N*-phthaloyl (Phth) groups. Introduction is by treatment of the primary amine with bis[2-(toluene-*p*-sulfonyloxy)ethyl]diphenylsilane (prepared in a three-step sequence from diphenyldichlorosilane) and $Et_3N$ in DMF at rt. Deprotection is with TBAF–CsF (1:1) in DMF or THF at rt. The group has a significant steric requirement, making protection of α-branched primary amines (*e.g.* most α-amino acids) slow.

A number of new carbamate PGs have been introduced this year. The prop-2-ynyloxycarbonyl (POC) group is removable with $(PhCH_2NEt_3)_2MoS_4$ and ultrasound[148] or $Co_2(CO)_8$ and TFA in $CH_2Cl_2$.[149] The former conditions allow removal from sulfur containing peptides. The triisopropylsilyloxycarbonyl (Tsoc) group is removable with TBAF in THF at 0 °C.[150] The 3,5-di-*tert*-butylbenzyloxycarbonyl group is a Cbz "replacement" with an improved solubility profile for polyurethane synthesis.[151] The 2-cyanoethoxycarbonyloxy (Ceoc) group has been used for protection of side-chain "aminolinkers" during oligonucleotide synthesis and is removable using $NH_4OH$.[152,153]

Two new transformations of primary ureas make their use as an amine PG a realistic possibility. Firstly, decarbamoylation can be effected quantitatively by treating solid monoalkylureas (including peptide derivatives) with gaseous $N_2O_4$ under an inert atmosphere.[154] Secondly, direct transformation to *N*-Boc derivatives can be achieved by treating the monoalkylureas with a copper(II) reagent prepared from $LiO^tBu$ and $CuBr_2$ in THF at rt.[155]

Sulfonamides are valuable amine derivatives largely due to their tendency towards crystallinity and resistance to nucleophiles and acids. However, deprotection remains a central issue in the use of this form of amine protection and so a new method for deprotection of primary phenyl and methyl sulfonamides in the presence of *N*-benzyl groups using TMSCl–NaI in refluxing MeCN over 3–4 h is significant.[156] Deprotection of secondary sulfonamides did not take place under these conditions. Deprotection of *N*-toluene-*p*-sulfonyl (Ts) groups from aziridines is particularly challenging as the aziridine ring is susceptible to ring-opening. A new method to accomplish this transformation employs sodium naphthalenide in DME at −78 °C.[157] Benzyl ethers survive this protocol although decomposition ensued on attempted deprotection of *N*-Ts-aziridine-2-carboxylic acid *tert*-butyl ester. The naphthalene-2-sulfonyl group has been introduced as an alternative to the Ts group as it is more readily deprotected using Mg in MeOH.[158] The group was selected following a detailed cyclic voltammetry study which identified this group as having a more favourable reduction potential. Selective deprotection in the presence of an *N*-Ts group is also possible. Primary sulfonamides are also useful as the acidic component in modified Mitsunobu type C–N coupling reactions. Best results when employing the 2,2,5,7,8-pentamethylchroman-6-sulfonyl (Pmc) group are obtained when employing the *N*,*N*,*N*′,*N*′-tetramethylazodicarboxamide (TMAD)–$PBu_3$ redox system.[159] The utility of the mercaptoethanol labile 2-nitrobenzenesulfonyl (Ns) group[160] in the synthesis of complex polyamines has been demonstrated in the synthesis of spider venom toxin HO-416b.[161]



HO-416b

43 T. Sano, K. Imai, K. Ohashi and T. Oriyama, *Chem. Lett.*, 1999, 265.
44 G. T. Copeland and S. J. Miller, *J. Am. Chem. Soc.*, 1999, **121**, 4306.
45 E. R. Jarvo, G. T. Copeland, N. Papaioannou, P. J. Bonitatebus and S. J. Miller, *J. Am. Chem. Soc.*, 1999, **121**, 11638.
46 F. Iwasaki, T. Maki, W. Nakashima, O. Onomura and Y. Matsumura, *Org. Lett.*, 1999, **1**, 969.
47 T. Sano, K. Ohashi and T. Oriyama, *Synthesis*, 1999, **7**, 1141.
48 P. Saravanan and V. K. Singh, *Tetrahedron Lett.*, 1999, **40**, 2611.
49 A. Orita, K. Sakamoto, Y. Hamada, A. Mitsutome and J. Otera, *Tetrahedron*, 1999, **55**, 2899.
50 P. Ilankumaran and J. G. Verkade, *J. Org. Chem.*, 1999, **64**, 9063.
51 K. K. Chauhan, C. G. Frost, I. Love and D. Waite, *Synlett*, 1999, 1743.
52 H. Gao and J. R. Dias, *Org. Prep. Proced. Int.*, 1999, **31**, 145.
53 A. Kabouche and Z. Kabouche, *Tetrahedron Lett.*, 1999, **40**, 2127.
54 S. Koeller and J.-P. Lellouche, *Tetrahedron Lett.*, 1999, **40**, 7043.
55 G. V. M. Sharma and A. Ilangovan, *Synlett*, 1999, 1963.
56 P. Ciuffreda, S. Casati and E. Santaniello, *Bioorg. Med. Chem. Lett.*, 1999, **9**, 1577.
57 R. Pfau and H. Kunz, *Synlett*, 1999, 1817.
58 L. Gardossi, R. Khan, P. A. Konowicz, L. Gropen and B. S. Paulsen, *J. Mol. Catal. B: Enzym.*, 1999, **6**, 89.
59 B. Yu, H. Yu, Y. Z. Hui and X. W. Han, *Synlett*, 1999, 753.
60 Y. Yoshida, Y. Sakakura, N. Aso, S. Okada and Y. Tanabe, *Tetrahedron*, 1999, **55**, 2183.
61 E. Bozó, S. Boros, J. Kuszmann and E. Gacs-Baitz, *Tetrahedron*, 1999, **55**, 8095.
62 G. Sabitha, S. Abraham, B. V. S. Reddy and J. S. Yadav, *Synlett*, 1999, 1745.
63 M. Curini, F. Epifano, M. C. Marcotullio, O. Rosati and U. Costantino, *Synth. Commun.*, 1999, **29**, 541.
64 S. Rover and P. Wipf, *Tetrahedron Lett.*, 1999, **40**, 5667.
65 A. Itoh, T. Kodama and Y. Masaki, *Synlett*, 1999, 357.
66 B. C. Ranu, U. Jana and A. Majee, *Tetrahedron Lett.*, 1999, **40**, 1985.
67 R. Hunter, W. Hinz and P. Richards, *Tetrahedron Lett.*, 1999, **40**, 3643.
68 K. P. R. Kartha and R. A. Field, *Synlett*, 1999, 311.
69 N. Tanaka and Y. Masaki, *Synlett*, 1999, 1960.
70 M. M. Mojtahedi, M. R. Saidi, M. M. Heravi and M. Bolourtchian, *Monatsh. Chem.*, 1999, **130**, 1175.
71 G. Sabitha, M. Syamala and J. S. Yadav, *Org. Lett.*, 1999, **1**, 1701.
72 R. D. Crouch, M. Stieff, J. L. Frie, A. B. Cadwallader and D. C. Bevis, *Tetrahedron Lett.*, 1999, **40**, 3133.
73 P. A. Grieco and C. J. Markworth, *Tetrahedron Lett.*, 1999, **40**, 665.
74 M. M. Heravi, D. Ajami and K. Tabar-Heydar, *Monatsh. Chem.*, 1999, **130**, 337.
75 M. M. Heravi, D. Ajami and M. Ghassemzadeh, *Synth. Commun.*, 1999, **29**, 781.
76 M. M. Heravi, D. Ajami and M. Ghassemzadeh, *Synthesis*, 1999, 393.
77 M. M. Heravi, M. Tajbakhsh and M. Ghassemzadeh, *Z. Naturforsch., Teil B: Chem. Sci.*, 1999, **54**, 394.
78 M. M. Heravi, D. Ajami and K. Tabar-Heydar, *Synth. Commun.*, 1999, **29**, 1009.
79 M. M. Mojtahedi, M. R. Saidi, M. Bolourtchian and M. M. Heravi, *Synth. Commun.*, 1999, **29**, 3283.
80 M. A. Brook, S. Balduzzi, M. Mohamed and C. Gottardo, *Tetrahedron*, 1999, **55**, 10027.
81 F. Chu, E. E. Dueno and K. W. Jung, *Tetrahedron Lett.*, 1999, **40**, 1847.
82 S. Buhler, H. Giegrich and W. Pfleiderer, *Nucleosides Nucleotides*, 1999, **18**, 1281.
83 A. Banerjee, K. Lee and D. E. Falvey, *Tetrahedron*, 1999, **55**, 12699.
84 M. H. Hyun, M. H. Kang and S. C. Han, *Tetrahedron Lett.*, 1999, **40**, 3435.
85 S. Majumdar and A. Bhattacharjya, *J. Org. Chem.*, 1999, **64**, 5682.
86 W.-L. Cheng, Y.-J. Shaw, S.-M. Yeh, P. P. Kanakamma, Y.-H. Chen, C. Chen, J.-C. Shieu, S.-J. Yiin, G.-H. Lee, Y. Wang and T.-Y. Luh, *J. Org. Chem.*, 1999, **64**, 532.
87 J. Zhang and M. D. Matteucci, *Tetrahedron Lett.*, 1999, **40**, 1467.
88 O. Arjona, F. Iradier, R. Medel and J. Plumet, *J. Org. Chem.*, 1999, **64**, 6090.
89 A. M. Kimbonguila, A. Merzouk, F. Guibe and A. Loffet, *Tetrahedron*, 1999, **55**, 6931.
90 P. Gomez-Martinez, A. M. Kimbonguila and F. Guibe, *Tetrahedron*, 1999, **55**, 6945.
91 T. Isobe and T. Ishikawa, *J. Org. Chem.*, 1999, **64**, 6984.
92 T. Isobe and T. Ishikaawa, *J. Org. Chem.*, 1999, **64**, 6989.
93 A. K. Chakraborti, A. B. Nandi and V. Grover, *J. Org. Chem.*, 1999, **64**, 8014.
94 B. Lohr, S. Orlich and H. Kunz, *Synlett*, 1999, 1136.
95 S. A. M. Merette, A. P. Burd and J. J. Deadman, *Tetrahedron Lett.*, 1999, **40**, 753.
96 H. Tamiaki, T. Obata and K. Toma, *Pept. Sci.*, 1999, **35**, 125.
97 M. J. Gaunt, C. E. Boschetti, J. Yu and J. B. Spencer, *Tetrahedron Lett.*, 1999, **40**, 1803.
98 A. Boutros, J.-Y. Legros and J.-C. Fiaud, *Tetrahedron Lett.*, 1999, **40**, 7329.
99 K. H. Cha, T. W. Kang, D. O. Cho, H.-W. Lee, J. Shin, K. Y. Jin, K.-W. Kim, J.-W. Kim and C.-I. Hong, *Synth. Commun.*, 1999, **29**, 3533.

157 S. C. Bergmeier and P. P. Seth, *Tetrahedron Lett.*, 1999, **40**, 6181.
158 B. Nyasse, L. Grehn, H. L. S. Maia, L. S. Monteiro and U. Ragnarsson, *J. Org. Chem.*, 1999, **64**, 7135.
159 A. S. Kolodziejczyk, E. Sugajska, B. Falkiewicz and K. Wisniewski, *Synlett*, 1999, 1606.
160 T. Fukuyama, M. Cheung and T. Kan, *Synlett*, 1999, 1301.
161 Y. Hidai, T. Kan and T. Fukuyama, *Tetrahedron Lett.*, 1999, **40**, 4711.
162 R. Lazny, J. Poplawski, J. Kobberling, D. Enders and S. Brase, *Synlett*, 1999, 1304.

US 20030092033A1

(19) United States

(12) Patent Application Publication

(10) Pub. No.: US 2003/0092033 A1

(43) Pub. Date: May 15, 2003

Weiner et al.

(54) METHODS FOR THE MANUFACTURE OF PURE SINGLE ENANTIOMER COMPOUNDS AND FOR SELECTING ENANTIOSELECTIVE ENZYMES

(76) Inventors: **David Weiner**, Del Mar, CA (US); **Tim Hitchman**, San Diego, CA (US); **Lishan Zhao**, Carlsbad, CA (US); **Mark J. Burk**, San Diego, CA (US); **Grace Desantis**, San Diego, CA (US); **Sarah Richardson Hanson**, San Diego, CA (US); **Aileen Milan**, San Diego, CA (US); **Toby Richardson**, San Diego, CA (US); **Patti Kretz**, San Marcos, CA (US); **Willaiam Greenberg**, San Diego, CA (US)

Correspondence Address:
**GREGORY P. EINHORN**
**Fish & Richardson P.C.**
**Suite 500**
**4350 La Jolla Village Drive**
**San Diego, CA 92122 (US)**

(21) Appl. No.: **10/178,938**

(22) Filed: **Jun. 21, 2002**

**Related U.S. Application Data**

(60) Provisional application No. 60/300,189, filed on Jun. 21, 2001. Provisional application No. 60/340,291, filed on Dec. 14, 2001.

**Publication Classification**

(51) **Int. Cl.**[7] ..... **C12Q 1/68**; G01N 33/53; C12N 9/16

(52) **U.S. Cl.** ..... **435/6**; 435/7.1; 435/196

(57) **ABSTRACT**

The invention provides biocatalytic methods for the manufacture of pure single enantiomer compounds. This invention provides methods of screening for enzymes which are highly enantioselective or enzymes that can provide any desired stereoisomer of a compound. The invention provides the use of single enantiomer substrates in performing a growth screen of a clonal library to identify highly stereoselective enzymes.

In one aspect, methods for screening and identification of enzymes, e.g., transaminases, nitrilases, aldolases, epoxide hydrolases are provided. Methods for the production and screening of gene libraries generated from nucleic acids isolated from more than one organism for enzyme, e.g., transaminase, activities are also provided.

60

30

**Figure 2**

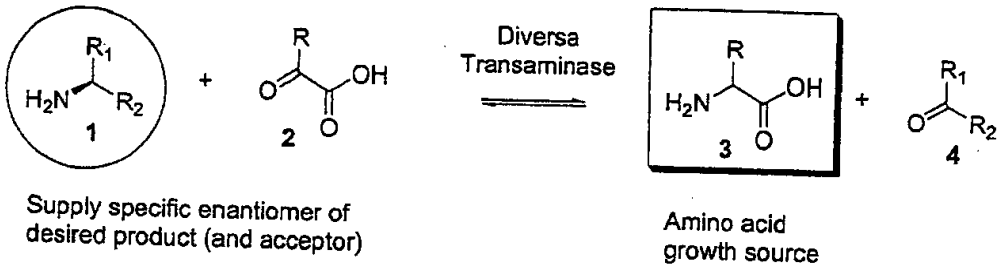

R1
H2N
R2
1
+
R
OH
O
2
Diversa
Transaminase
R
H2N
OH
O
3
+
R1
O
R2
4
Supply specific enantiomer of
desired product (and acceptor)
Amino acid
growth source

Figure 4

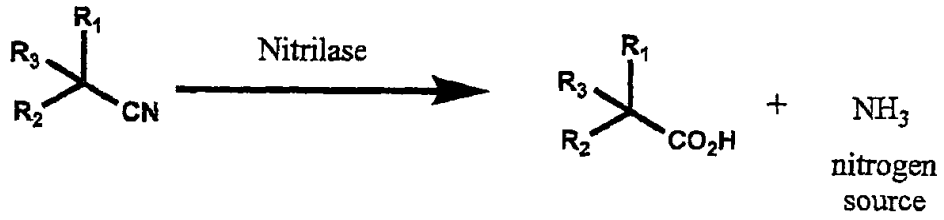
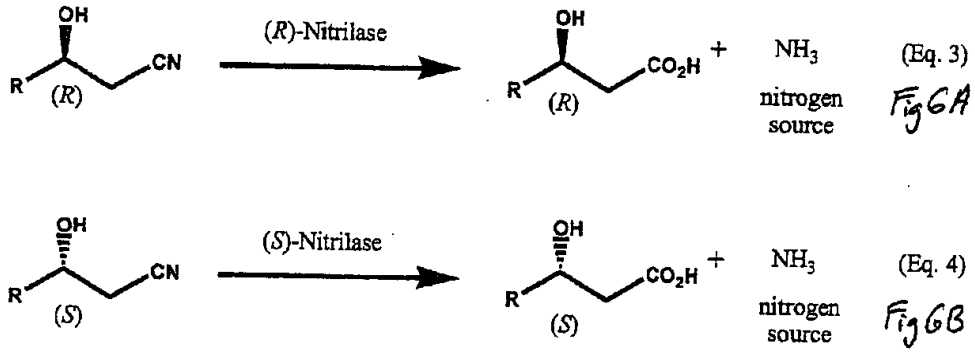


Figure 6


OH
CN
R
(R)
(R)-Nitrilase
OH
CO2H
R
(R)
+
NH3
nitrogen source
(Eq. 3)
Fig 6A
OH
CN
R
(S)
(S)-Nitrilase
OH
CO2H
R
(S)
+
NH3
nitrogen source
(Eq. 4)
Fig 6B

## Figure 8

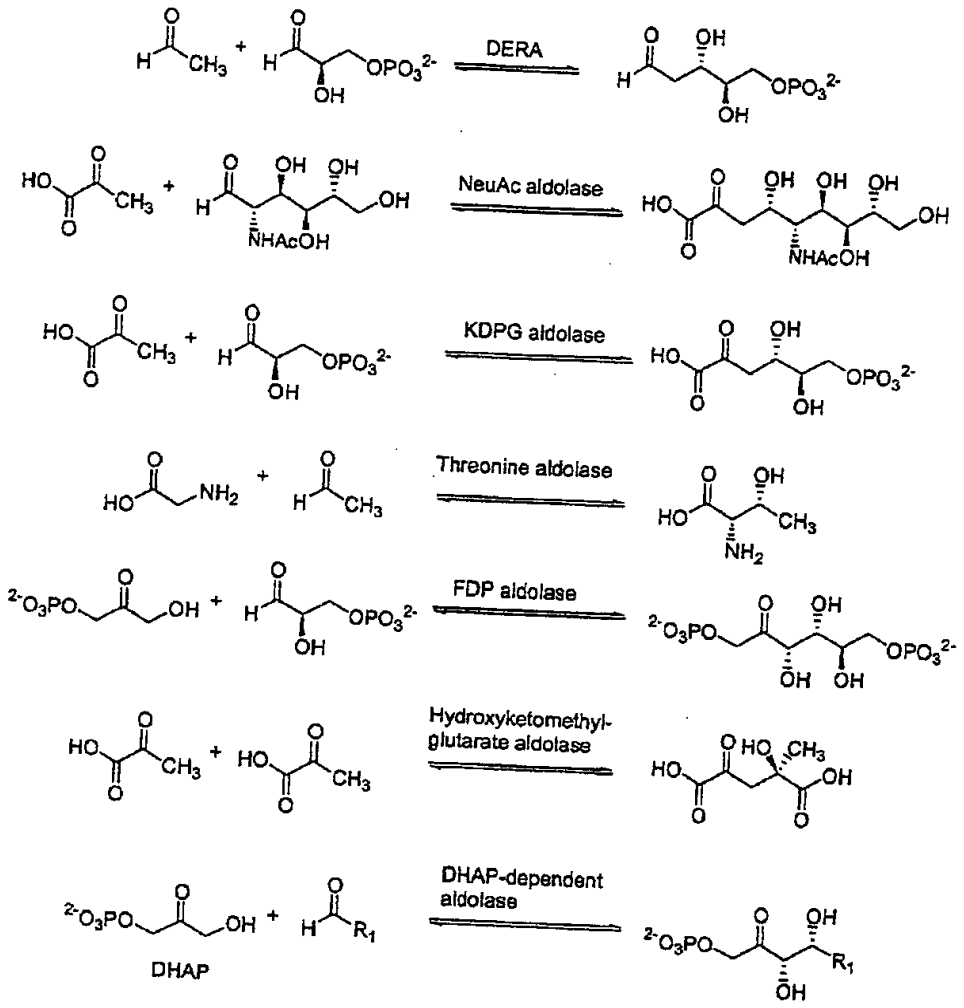

DERA
NeuAc aldolase
KDPG aldolase
Threonine aldolase
FDP aldolase
Hydroxyketomethyl-glutarate aldolase
DHAP-dependent aldolase
DHAP

**Figure 10**

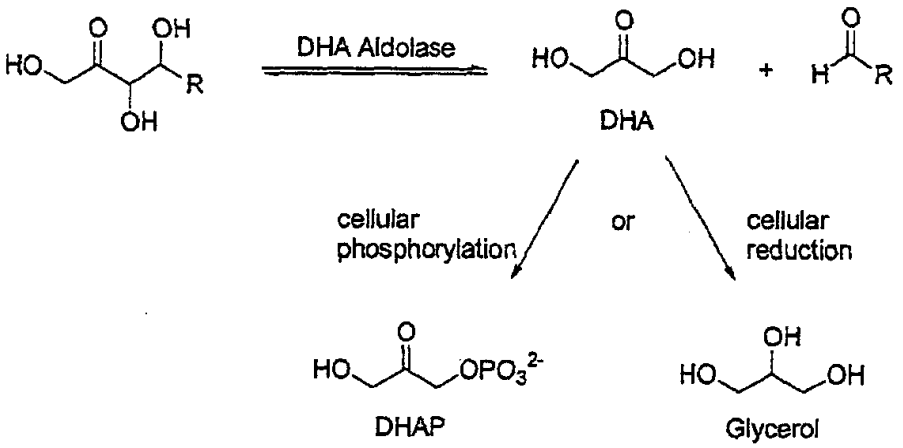

DHA Aldolase
DHA
cellular phosphorylation
or
cellular reduction
DHAP
Glycerol

**Figure 12**

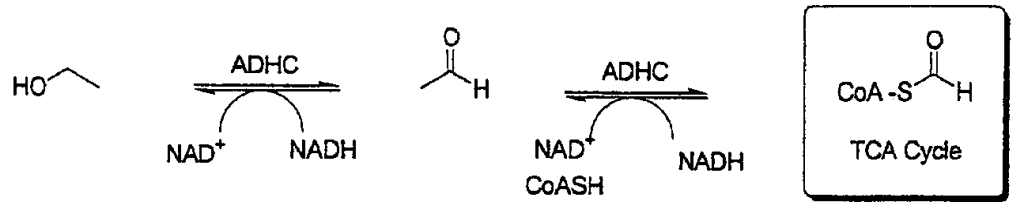


**Figure 14**

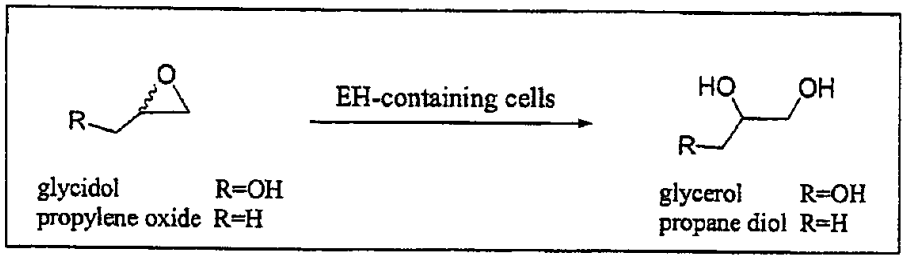
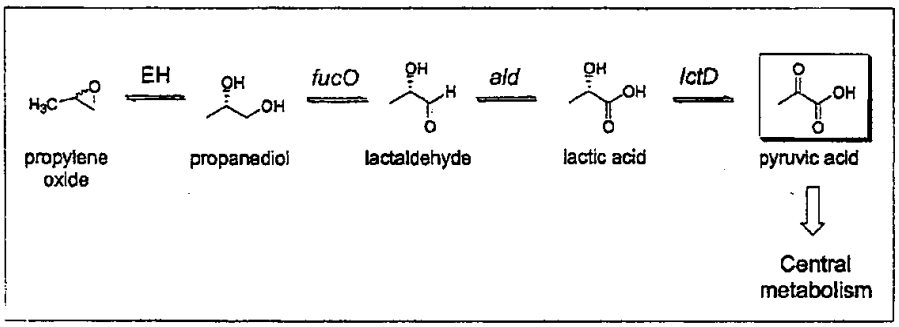

O
R
EH-containing cells
HO
OH
R
glycidol R=OH
propylene oxide R=H
glycerol R=OH
propane diol R=H

Figure 16


EH
fucO
ald
lctD
propylene oxide
propanediol
lactaldehyde
lactic acid
pyruvic acid
Central metabolism

substrates, an amino compound (amino donor) and a keto compound (amino acceptor). The transaminase catalyzes the exchange of the keto group from the keto compound and the amino group from the amino compound. This exchange generates a new amino compound from the keto compound and a new keto compound from the amino compound. Typically only one of the products is desired, generally the new amino compound, and the other is an unwanted by-product. Used in isolation, the enzyme converts the two substrates to the two products. Theoretically, because the reaction is reversible, it proceeds until it reaches equilibrium.

[0012] U.S. Pat. No. 4,518,692 ("Rozzell I") discloses a method for producing L-amino acids by reacting L-aspartic acid and various 2-keto acids with transaminases. The Rozzell I method uses L-aspartic acid as the amino acid to produce oxaloacetate and describes various methods of decarboxylating oxaloacetate to form pyruvate. However, the pyruvate produced in the Rozzell I method can still act as a keto donor in the reverse process to form alanine. Tokarski et al., Biotechnology Letters, Vol. 10 (1) (1988), pp. 7-10, show that alanine acts as a substrate in transaminase reactions. See also, Transaminases (1985); and Amino Acids: Biosynthesis and Genetic Regulation, Klaus M. Herrmann and Ronald L. Somerville ed. (1983) (Addison-Wesley Publishing, Reading Mass.). Tokarski, et al. studied the use of a transaminase to produce L-2-aminobutyrate from 2-ketobutyrate and alanine. The reference, however, discloses only 25-30% conversion to products, demonstrating that the reverse reaction is very significant. This has long been considered an intrinsic property and a problem of transaminase reactions and is the major reason such enzyme catalyzed reactions have not been more often exploited in industrial processes to produce these highly desired amine products.

[0013] U.S. Pat. No. 4,826,766 ("Rozzell II") discloses an improved transaminase catalyzed reaction that employs two transaminase enzymes and additional keto acids. In the process, a first transaminase enzyme catalyzes the reaction between a first amino acid and a first keto acid to produce a second amino acid and second keto acid. A second transaminase enzyme then catalyzes a further reaction of the second amino acid and a third keto acid to form the desired amino acid. The two transaminase enzymes are selected such that the first enzyme does not catalyze the second reaction and the second enzyme does not catalyze the first reaction.

[0014] Another transaminase process, which combines the transaminase enzyme with a second enzyme that eliminates the keto acid produced by the transaminase reaction, preventing the attainment of equilibrium, and driving the amino acid producing reaction to completion, is known from U.S. Pat. No. 6,197,558 (Fotheringham). The second enzyme catalyzes a reaction, which converts the keto acid to a substance that can no longer react with the transaminase. By removing the second keto acid, the second enzyme allows the amino acid producing reaction to proceed to an extent that the desired amino acid product represents approximately 100% of the amino acids produced.

[0015] Aldolases are ubiquitous enzymes that catalyze the formation of carbon-carbon bonds through the aldol reaction (**FIG. 7**). Depending on the donors and acceptors utilized, the reaction generates one or two new stereocenters (indicated by the asterixes in **FIG. 1**). Thus, aldolases have great potential for the production of advanced chiral products that are difficult and/or expensive to produce by traditional chemical routes. **FIG. 8** illustrates a few well-characterized examples of reactions catalyzed by aldolases. In regard to substrate specificity, the aldehyde acceptor component can be varied to some extent (**FIG. 3**), and the enolate donor requirement is typically quite strict. Some examples of aldolase-mediated synthesis with non-natural substrates have been reported, although these cases are currently limited (see, e.g., JOC 2000, 95, 8264; b. JACS 1996, 118, 2117; c. JACS 1997, 119, 11734). Realization of the synthetic potential of aldolases in large scale industrial processes has been limited by the lack of available enzymes with the necessary properties.

[0016] There are two major routes from a nitrile to an analogous acid: (1) a nitrilase catalyzes the direct hydrolysis of a nitrile to a carboxylic acid with the concomitant release of ammonia; or (2) a nitrile hydratase adds a molecule of water across the carbon-nitrogen bonding system to give the corresponding amide, which then acts as a substrate for an amidase enzyme which hydrolyzes the carbon-nitrogen bond to give the carboxylic acid product with the concomitant release of ammonia. The nitrilase enzyme therefore provides the more direct route to the acid.

[0017] A nitrile group offers many advantages in devising synthetic routes in that it is often easily introduced into a molecular structure and can be carried through many processes as a masked acid or amide group. This is only of use, however, if the nitrile can be unmasked at the relevant step in the synthesis. Cyanide represents a widely applicable $C_1$-synthon (cyanide is one of the few water-stable carbanions) which can be employed for the synthesis of a carbon framework. However, further transformations of the nitrile thus obtained are impeded due to the harsh reaction conditions required for its hydrolysis using normal chemical synthesis procedures. The use of enzymes to catalyze the reactions of nitrites is attractive because nitrilase enzymes are able to effect reactions with fewer environmentally hazardous reagents and by-products than in many traditional chemical methods. Indeed, the chemoselective biocatalytic hydrolysis of nitrites represents a valuable alternative because it occurs at ambient temperature and near physiological pH.

[0018] The importance of asymmetric organic synthesis in drug design and discovery has fueled the search for new synthetic methods and chiral precursors which can be utilized in developing complex molecules of biological interest. One important class of chiral molecules are the α-substituted carboxylic acids, which include the α-amino acids. These molecules have long been recognized as important chiral precursors to a wide variety of complex biologically active molecules, and a great deal of research effort has been dedicated to the development of methods for the synthesis of enantiomerically pure α-amino acids and chiral medicines.

[0019] Of particular use to synthetic chemists who make chiral medicines would be an enzyme system which is useful under non-sterile conditions, which is useful in non-biological laboratories, which is available in a form convenient for storage and use; which has broad substrate specificity, which acts on poorly water soluble substrates; which has predict-

[0026] All of the reactions of the invention, whether cell-based or in vitro, can take place entirely or partly in an array, e.g., a double-orificed capillary array, such as a GIGAMATRIX™ capillary array.

[0027] The invention provides a method for identifying a nucleic acid encoding an enantioselective enzyme comprising the following steps: (a) providing a nucleic acid library; (b) providing a precursor of a specific chirality for a composition essential for growth, wherein the precursor is capable of being enzymatically converted to a product comprising the composition essential for growth, and to be growth-stimulating the composition essential for growth must have a chirality corresponding to the chirality of the precursor; (c) providing a plurality of cells, wherein the cells cannot make the composition essential for growth; (d) inserting in a cell a member of the gene library and culturing the cells in a medium lacking the composition essential for growth; (e) adding the precursor of step (b) to the culture; (f) selecting a growing cell and identifying the inserted library member of step (d), wherein the cell is capable of growth by enzymatic conversion of the precursor to a product comprising the composition essential for growth, and the enzyme is encoded by the library member, thereby identifying a nucleic acid encoding an enantioselective enzyme. The precursor can be added to the cells or medium at any point. Alternatively, the precursor can be endogenously produced by the cells.

[0028] The invention provides a growth selection screen using single enantiomer substrates to discover enzymes with a specific stereoselectivity profile comprising the following steps: (a) providing a nucleic acid or a polypeptide library; (b) providing a single enantiomer substrate for a composition essential for growth, wherein the substrate is capable of being converted to a product comprising a composition essential for growth, and to be growth-stimulating the composition essential for growth must have a chirality corresponding to the chirality of the precursor; (c) providing a plurality of cells, wherein the cells cannot make the composition essential for growth; (d) inserting in the cells a member of the nucleic acid or polypeptide library and culturing the cells in a medium lacking the composition essential for growth; (e) adding the single enantiomer substrate of step (b) to the culture; and (f) selecting a growing cell and identifying the inserted nucleic acid or polypeptide of step (d), wherein the cell is capable of growth by enzymatic conversion of the single enantiomer substrate to a product comprising the composition essential for growth, thereby identifying an enzyme with a specific stereoselectivity profile.

[0029] In one aspect, the nucleic acid is a member of a nucleic acid or gene library, e.g., a DNA or cDNA library. The library can be obtained from a pure culture or a mixed population of organisms. The mixed population of organisms can be derived from an environmental sample, e.g., a soil sample, a water sample or an air sample.

[0030] The methods of the invention can identify a coding sequence for an enzyme with a specific stereoselectivity profile or a polypeptide with a specific stereoselectivity profile for any and all enzymes, e.g., a transaminase, a nitrilase, an aldolase or a hydrolase, e.g., an epoxide hydrolase, a protease, a lipase and the like.

[0031] In practicing the invention, the equilibrium of the reaction can be manipulated by any means, e.g., adding substrate/ precursor or removing reaction product or both. For example, if a reaction is in equilibrium, a product can be removed to drive reaction toward the product side and/or a substrate can be removed to drive the reaction to substrate side. In one aspect, the equilibrium of the conversion of the substrate or precursor to the product is shifted in the direction of product formation by addition of an excess of substrate or precursor. Alternatively, the equilibrium of the conversion of the product to the substrate or precursor is shifted in the direction of substrate or precursor formation by addition of an excess of product.

[0032] In one aspect, the enzyme is a transaminase and the substrate or precursor is a specific enantiomer of an amino acid and the product is a specific enantiomer of an amino donor. The product can further comprises an α-keto acid. The substrate or precursor can comprises a specific enantiomer of an amino donor and the product is a specific enantiomer of an amino acid. The equilibrium of the conversion can be shifted in the direction of amino acid product formation by addition of excess amino donor. The method can further comprise adding an α-keto acid amino acceptor to the media. The equilibrium of the conversion of the product to the substrate or precursor can be shifted in the direction of product formation by enzymatic removal of an α-keto acid product. The equilibrium of the conversion of the specific enantiomer to the specific growth source can shifted in the direction of product formation by chemical removal of α-keto acid product.

[0033] In one aspect, the enzyme is a nitrilase, and by supplying only nitrile groups of a desired chirality enantioselective nitrilase enzymes are identified. The substrate or precursor can comprise a specific enantiomer of a nitrile-containing compound and the product comprises a specific enantiomer of a corresponding carboxylic acid and ammonia. The substrate or precursor can comprise a specific enantiomer of a carboxylic acid and the product comprises a specific enantiomer of a nitrile-containing compound. The medium can comprise a nitrogen-free minimal media for cell growth, thereby only clones that can hydrolyze a nitrile group will produce the nitrogen source required to grow.

[0034] In one aspect, the enzyme is an aldolase and deoxyribose-5-phosphate comprises the substrate or precursor and acetaldehyde comprises a product of the reaction. The product can further comprise a glyceraldehyde-3-phosphate. The enzyme can be an aldolase and a deoxyribose, a 5-O-methyl-deoxyribose, or a dideoxyribose can comprise the substrate or precursor and an acetaldehyde can comprise a product of the reaction. The precursor or substrate can comprise a deoxyribose and the product can further comprise a glyceraldehyde. The precursor or substrate can comprise a 5-O-methyl-deoxyribose and the product can further comprise a 3-O-methyl-deoxyribose. The precursor or substrate can comprise a dideoxyribose and the product can further comprise a lactaldehyde.

[0035] In one aspect, the nucleic acid further comprises an expression cassette, an expression vector, a phage or a plasmid. The vector can be a PAC, a BAC, a MAC or a YAC. The nucleic acid library can comprise phagemid library cells.

[0036] In one aspect, the medium comprises a solid substrate or a liquid media.

[0037] In one aspect, the cells that cannot make a factor, element or composition essential for growth are auxotrophs.

consuming methods for characterization and identification or were unable to be identified or characterized.

[0048] In one aspect, the invention provides a method for obtaining a bioactivity or a biomolecule of interest by screening a library of clones generated from nucleic acids from a mixed population of cells, for a specified bioactivity or biomolecule, variegating a nucleic acid sequence contained in a clone having the specified bioactivity or biomolecule; and comparing the variegated bioactivity or biomolecule with the specified bioactivity or biomolecule wherein a difference in the bioactivity or biomolecule is indicative of an effect of sequence variegation, thereby providing the bioactivity or biomolecule of interest.

[0049] In another aspect, the invention provides a method for identifying a bioactivity or a biomolecule of interest by screening a library of clones generated from pooled nucleic acids obtained from a plurality of isolates for a specified bioactivity or biomolecule; and identifying a clone which contains the specified bioactivity or biomolecule.

[0050] In yet another aspect, the invention provides a method for identifying a bioactivity or a biomolecule of interest. The method includes screening a library of clones generated from pooled nucleic acids obtained from a plurality of isolates for a specified bioactivity or biomolecule, variegating a nucleic acid sequence contained in a clone having the specified bioactivity or biomolecule, and comparing the variegated bioactivity or biomolecule with the specified bioactivity or biomolecule wherein a difference in the bioactivity or biomolecule is indicative of an effect of introducing at least one sequence variegation, thereby providing the bioactivity or biomolecule of interest.

[0051] In another aspect, the invention provides a method for identifying a bioactivity or a biomolecule of interest, wherein the method includes screening a library of clones generated from pooling individual gene libraries generated from the nucleic acids obtained from each of a plurality of isolates for a specified bioactivity or biomolecule and identifying a clone which contains the specified bioactivity or biomolecule.

[0052] In another aspect, the invention provides a method for identifying a bioactivity or a biomolecule of interest by screening a library for a specified bioactivity or biomolecule wherein the library is generated from pooling individual gene libraries generated from the nucleic acids obtained from each of a plurality of isolates, variegating a nucleic acid sequence contained in a clone having the specified bioactivity or biomolecule, and comparing the variegated bioactivity or biomolecule with the specified bioactivity or biomolecule wherein a difference in the bioactivity or biomolecule is indicative of an effect of introducing at least one sequence variegation, thereby providing the bioactivity or biomolecule of interest.

[0053] In one aspect, the invention provides a method of identifying a bioactivity or biomolecule of interest, including screening a library of clones generated from the nucleic acids from an enriched population of organisms for a specified bioactivity or biomolecule and identifying a clone containing the specified bioactivity or biomolecule.

[0054] In one aspect, the invention provides a method of identifying a bioactivity or biomolecule of interest by screening a library of clones generated from nucleic acids from an enriched population of organisms for a specified bioactivity or biomolecule, variegating a nucleic acid sequence contained in a clone having the specified bioactivity or biomolecule, and comparing the variegated bioactivity or biomolecule with the specified bioactivity or biomolecule wherein a difference in the bioactivity or biomolecule is indicative of an effect of introducing at least one sequence variegation, thereby providing the bioactivity or biomolecule of interest.

[0055] In one aspect, the invention provides a method for identifying a bioactivity or a biomolecule of interest. The bioactivity or biomolecule of interest is identified by incubating nucleic acids from a mixed population of organisms with at least one oligonucleotide probe having a detectable molecule and at least a portion of a nucleic acid sequence encoding a molecule of interest under conditions to allow interaction of complementary sequences, identifying nucleic acid sequences having a complement to the oligonucleotide probe using an analyzer that detects the detectable molecule. A library is then generated from the identified nucleic acid sequences and the library is screened for a specified bioactivity or biomolecule. Nucleic acid sequence contained in a clone having the specified bioactivity or biomolecule is variegated and the variegated bioactivity or biomolecule compared with the specified bioactivity or biomolecule wherein a difference in the bioactivity or biomolecule is indicative of an effect of introducing at least one sequence variation, thereby providing the bioactivity or biomolecule of interest.

[0056] In one aspect, the invention provides a method for identifying a bioactivity or a biomolecule of interest by co-encapsulating in a microenvironment nucleic acids obtained from a mixed population of organisms, with at least one oligonucleotide probe having a detectable molecule and at least a portion of a nucleic acid sequence encoding a molecule of interest under such conditions and for such time as to allow interaction of complementary sequences, identifying encapsulated nucleic acids containing a complement to the oligonucleotide probe encoding the molecule of interest by separating the encapsulated nucleic acids with an analyzer that detects the detectable molecule, generating a library from the separated encapsulated nucleic acids, screening the library for a specified bioactivity or biomolecule, variegating a nucleic acid sequence contained in a clone having the specified bioactivity or biomolecule, and comparing the variegated bioactivity or biomolecule with the specified bioactivity or biomolecule wherein a difference in the bioactivity or biomolecule is indicative of an effect of introducing at least one sequence variation, thereby providing the bioactivity or biomolecule of interest.

[0057] In one aspect, the invention provides a method including co-encapsulating in a microenvironment nucleic acids obtained from an isolate of a mixed population of organisms, with at least one oligonucleotide probe having a detectable marker and at least a portion of a polynucleotide sequence encoding a molecule having a bioactivity of interest under conditions and for such time as to allow interaction of complementary sequences, identifying encapsulated nucleic acids containing a complement to the oligonucleotide probe encoding the molecule of interest by separating the encapsulated nucleic acids with an analyzer that detects the detectable marker, generating a library from the separated encapsulated nucleic acids, screening the library for a

## DETAILED DESCRIPTION

[0079] The present invention provides methods for making compounds, e.g., amino acids, of a specific chirality, i.e., the product of the method is a single enantiomer. Also provided are methods for selecting enzymes whose reaction products are of a specific chirality, i.e., the product of the enzyme's reaction is a single enantiomer.

[0080] The present invention presents methods for discovery and identification of enzymes that are highly enantioselective and enzymes that can provide a desired stereoisomer of a given chiral compound. In one aspect, the methods comprise performing a growth selection screen on a clonal library whereby single enantiomer substrates are used to discover enzymes with a specific stereoselectivity profile.

[0081] In one exemplary method of the invention, a growth selection screen is performed on a clonal library to generate a single enantiomer reaction product and to discover enzymes with a specific stereoselectivity profile. For example, the methods provides a selection strategy that uses the reversibility of enzyme-catalyzed reactions to advantage and, in essence, searches for clones that can run a reaction backwards from a desired substrate of a specific chirality, e.g., an amine, to produce a desired growth source, e.g., an amino acid, of a specific chirality.

[0082] For example, the invention provides a selection strategy that uses the reversibility of transaminase-catalyzed reactions to advantage and, in essence, searches for clones that can run a reaction backwards from a desired amine to produce a growth source, i.e., a composition essential for growth of the cell, such as an amino acid. The exemplary method comprises:

[0083] Phagemid library cells are inoculated into media lacking a crucial growth source 3 (see **FIG. 2**). Both solid and liquid selections can be used.

[0084] Any desired amine (or amino acid) 1 (see **FIG. 2**) is added to the media. Depending on the natural cellular background levels of the amine acceptor 2 (see **FIG. 2**), it may or may not be necessary to add exogenous 2.

[0085] Only clones harboring an active transaminase that can convert 1 to 3 (see **FIG. 2**) will grow. See **FIGS. 2 and 3**. Any host strain can be used. Any desired host strain can be made using routine selection methods. For example, several amino acid auxotrophs (e.g. aspartate or glutamate) can be made via knockout strategies, e.g., transposon mutagenesis.

[0086] The selection strategies of the invention are powerful because they offer an extremely high throughput method for screening environmental DNA libraries for enzyme-encoding genes, e.g., transaminase or nitrilases genes. In one aspect, the desired product of a particular biotransformation is used to discover the enzyme for its own synthesis. Thus, the selection is highly specific. Extremely rare biocatalysts are accessible via the methods of the invention. Any chiral amine or amino acid can be used. In one aspect, chiral amines or amino acids that are non-toxic to the cell and are able to permeate the cell membrane are used.

[0087] The methods of the invention can use any biocatalytic or metabolic pathway manipulation scheme or methodology, and, the methods of the invention can use any culturing or selection strategy, see, e.g., Hoch, et al., WO 00/22170, U.S. Patent Application Publication Nos. 20020050476; 20020012974; 20020006644.

[0088] In another exemplary method, the invention provides growth selection with nitrilases. Nitrilases are a class of enzymes that catalyze the hydrolysis of nitrile-containing compounds into the corresponding carboxylic acid and ammonia. By using nitrogen-free minimal media for growth, only clones that can hydrolyze a nitrile group will produce the nitrogen source required to thrive. In one aspect, by supplying only nitrile groups of a desired chirality, enantioselective enzymes can be identified. By utilizing single enantiomer substrates for the growth selection and identification of enzymes, clones in a library that will rapidly hydrolyze that enantiomer of the substrate can be selected.

[0089] In another aspect, opposite enantiomers of a given product can also be selected. By employing the opposite enantiomer of a substrate, only enzymes that can readily hydrolyze the opposite enantiomer will grow. As discussed in Example 1, this exemplary method entails utilizing single enantiomer nitrites. **FIG. 5A** (Eq. 1) shows that clones possessing an active nitrilase enzyme (Nitrilase-1) that can hydrolyze the nitrile enantiomer shown will grow effectively. By corollary, only clones possessing Nitrilase-2 will survive if provided with the opposite nitrile enantiomer, as shown in **FIG. 5B** (Eq. 2).

[0090] This selection scheme may allow discovery of two different enzymes which may provide either enantiomer of a desired product. The invention also can be applied to other types of nitrites as well as many other classes of enzymes and substrates.

[0091] The present invention provides methods of selecting and creating a wide range of transaminases. Transaminase catalyzed reactions are reversible and the equilibrium constant is generally near unity. Even though it is a challenge to obtain high yields of products several methods as described above have been developed to shift the equilibrium towards product formation direction. The method of selecting trans aminases of the present invention utilizes the reversibility of transaminase-catalyzed reactions to its advantage and, in essence, searches for clones that can run a reaction backwards from a desired amine (or amino acid) to produce a growth source. See **FIGS. 2 and 3**. Once the clones are identified, the desired transaminases may be isolated from the clones.

[0092] In one aspect, the method of selecting transaminases according to the present invention includes the steps of:

[0093] Inoculating clones containing phagemid library cells into media lacking a crucial growth source 3 of **FIG. 2**. Both solid and liquid selections can be used.

[0094] Adding a particular desired amine or amino acid 1 of **FIG. 2** to the media. Depending on the natural cellular background levels of the amine acceptor 2 of **FIG. 2**, it may or may not be necessary to add exogenous amine acceptor 2 of **FIG. 2**.

[0095] Detecting the clones having substantial cell growths to determine the existence of transaminases because only clones harboring an active transaminase that can convert amine or amino acid 1 to growth source 3 will grow.

reaction mixture to the desired amino acid product. The cells may be permeabilized to facilitate diffusion of the substrates and products into and out of the cells. This permeabilization can be accomplished by treating cells with a low concentration of a surfactant, including but not limited to TWEEN 80, TRITON X-100, NONIDET P40, cetylpyridinium chloride, deoxycholic acid, hexadecyltrimethylammonium bromide or benzalkonium chloride. Further, organic solvents, including but not limited to N,N-dimethylformamide (DMF), dimethyl sulfoxide (DMSO), ethanol or acetone at low concentrations have also been used to increase permeabilization.

[0110] Transaminases may also be added to the starting reaction mixture in the form of cell extracts containing crude, partially purified, or purified enzyme. Cell extracts are prepared by methods known to those skilled in the art, which provide for cell disruption and recovery of the enzyme. Cell disruption, can be accomplished by mechanical or non-mechanical means. Most often, for bacterial suspensions mechanical devices such as a French pressure cell, ultrasonication, bead mill or Manton-Gaulin homogenizer is used with the specifics of the method known to those of ordinary skill in the art. See, Scopes, R. K. "Protein Purification", (1982) (Springer-Verlag, New York). The reaction using the cell extract is then carried out in similar fashion to the whole cell method discussed above.

[0111] The enzyme-containing cells, or extracts thereof or purified enzyme or enzyme fractions, may also be immobilized, if desired. Immobilization methods, which may be used in the practice of this invention include well-known methods such as entrapment in polymeric gels, covalent attachment, crosslinking, adsorption, and encapsulation. Some examples of these methods are described by A. M. Klibanov in Science, 219:722-727 (1983) and the references therein and in Methods in Enzymology (1976), Volume 44, (K. Mosbach editor) which are hereby incorporated by reference.

[0112] In one method of immobilization disclosed in U.S. Pat. No. 5,019,509, a support material containing at least 20% by weight of silica or alumina is contacted with aminoalkyl compound such as an aminoalkyl silane, polyethyleneimine, or a polyalkylamine, followed by activation with glutaralde-hyde. The enzyme-containing solution is then contacted with the activated support to produce an immobilized enzyme composition having transaminase- and/or acetolactate synthase activity. Other immobilization supports useful in the practice of this invention include, but are not limited to, porous glass and porous ceramics, bentonite, diatomaceous earth, charcoal SEPHAROSE® and SEPHAROSE® derivatives, cellulose and cellulose derivatives, polyacrylamide and polyacrylamide derivatives, polyazetidine, alginate, carrageenan, and CHROMOSORB®. SEPHAROSE® (Pharmacia Fine Chemicals, Uppsala Sweden) is a bead-formed gel prepared from agarose. The manufacturer's product literature reports that in its natural state, agarose occurs as part of the complex mixture of charged and neutral polysaccharides referred to as agar. The agarose used to make SEPHAROSE® is obtained by a purification process which removes the charged polysaccharides to give a gel with only a very small number of residual charged groups. Those of ordinary skill in the art will appreciate that a number of other materials suitable for the immobilization of cells or extracts derived therefrom may also be useful for the immobilization of the enzymes used in the present invention. These supports can be activated, if desired, by techniques well-known in the art.

[0113] The selection process to produce a desired growth source is carried out by contacting a solution containing a first keto acid and a first amino acid with the enzymes under conditions permitting the conversion of at least a portion of the first keto acid to the desired amino acid. In the practice of the processes of this invention the cells contact an aqueous solution of the enzymes at a cell concentration in the range of about 50 mg/ml to about 200 mg/ml. In one aspect the cell concentration is about 100 mg/ml. When the invention is practiced using extracts of cells, the extracts are prepared from an amount of cells that would give these cell concentrations.

[0114] The enzymatic reactions of this invention are carried out at temperatures in the range of from about 30° C. to about 50° C., and preferably at temperatures ranging from about 37° C. to about 45° C. The optimal pH for the reaction ranges from about 6 to about 9, and more preferably from about 7 to about 8, with a pH of 8 being most preferred.

[0115] The present invention also includes recombinant constructs comprising one or more of the sequences encoding transaminases selected by the selection method of the present invention. The constructs comprise a vector, such as a plasmid or viral vector, into which a sequence of a transaminase selected by the invention has been inserted, in a forward or reverse orientation. In one aspect of this aspect, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example; Bacterial: pQE70, pQE60, pQE-9 (Qiagen), pBluescript II KS, ptrc99a, pKK223-3, pDR540, pRIT2T (Pharmacia); Eukaryotic: pXT1, pSG5 (Stratagene) pSVK3, pBPV, pMSG, pSVL SV40 (Pharmacia). However, any other plasmid or vector may be used as long as they are replicable and viable in the host.

[0116] Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are pKK232-8 and pCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P.sub.R, P.sub.L and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

[0117] In a further aspect, the present invention relates to host cells containing the above-described constructs. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-Dextran mediated transfection, or electroporation (Davis, L., Dibner, M., Battey, I., Basic Methods in Molecular Biology, (1986)).

[0118] The constructs in host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence. Alternatively, the transaminases can be synthetically produced by conventional peptide synthesizers.

[0131] In addition, fluorescence activated cell sorting can be used to screen and isolate clones having an activity or sequence of interest. Previously, FACS machines have been employed in the studies focused on the analyses of eukaryotic and prokaryotic cell lines and cell culture processes. FACS has also been utilized to monitor production of foreign proteins in both eukaryotes and prokaryotes to study, for example, differential gene expression, and the like. The detection and counting capabilities of the FACS system have been applied in these examples. However, FACS has never previously been employed in a discovery process to screen for and recover bioactivities in prokaryotes. Furthermore, the present invention does not require cells to survive, as do previously described technologies, since the desired nucleic acid (recombinant clones) can be obtained from alive or dead cells. The cells only need to be viable long enough to produce the compound to be detected, and can thereafter be either viable or non-viable cells so long as the expressed biomolecule remains active. The present invention also solves problems that would have been associated with detection and sorting of *E. coli* expressing recombinant enzymes, and recovering encoding nucleic acids. Additionally, the present invention includes within its aspects any apparatus capable of detecting fluorescent wavelengths associated with biological material, such apparati are defmed herein as fluorescent analyzers (one example of which is a FACS apparatus).

[0132] In one aspect, the invention identifies nucleic acid sequences from a mixed population of organisms, isolates, or enriched populations. In this aspect, it is not necessary to express gene products. Nucleic acid sequences of interest can be identified or "biopanned" by contacting a clone, device (e.g. a gene chip), filter, or nucleic acid sample with a probe labeled with a detectable molecule. The probe will typically have a sequence that is substantially identical to the nucleic acid sequence of interest. Alternatively, the probe will be a fragment or full-length nucleic acid sequence encoding a polypeptide of interest. The probe and nucleic acids are incubated under conditions and for such time as to allow the probe and a substantially complementary sequence to hybridize. Hybridization stringency will vary depending on, for example, the length and GC content of the probe. Such factors can be determined empirically (See, for example, Sambrook et al., Molecular Cloning—A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1989, and Current Protocols in Molecular Biology, M. Ausubel et al., eds., (Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., most recent Supplement)). Once hybridized the complementary sequence can be PCR amplified, identified by hybridization techniques (e.g., exposing the probe and nucleic acid mixture to a film), or detecting the nucleic acid using a chip.

[0133] Once a sequence or bioactivity of interest is identified (e.g., an enzyme of interest) the sequence or polynucleotide encoding the bioactivity of interest can be evolved, mutated or derived to modify the amino acid sequence to provide, for example, modified activities such as increased thermostability, specificity or activity.

[0134] An "amino acid" is a molecule having the structure wherein a central carbon atom (the α-carbon atom) is linked to a hydrogen atom, a carboxylic acid group (the carbon atom of which is referred to herein as a "carboxyl carbon atom"), an amino group (the nitrogen atom of which is referred to herein as an "amino nitrogen atom"), and a side chain group, R. When incorporated into a peptide, polypeptide, or protein, an amino acid loses one or more atoms of its amino acid carboxylic groups in the dehydration reaction that links one amino acid to another. As a result, when incorporated into a protein, an amino acid is referred to as an "amino acid residue."

[0135] "Protein" or "polypeptide" refers to any polymer of two or more individual amino acids (whether or not naturally occurring) linked via a peptide bond, and occurs when the carboxyl carbon atom of the carboxylic acid group bonded to the α-carbon of one amino acid (or amino acid residue) becomes covalently bound to the amino nitrogen atom of amino group bonded to the α-carbon of an adjacent amino acid. The term "protein" is understood to include the terms "polypeptide" and "peptide" (which, at times may be used interchangeably herein) within its meaning. In addition, proteins comprising multiple polypeptide subunits (e.g., DNA polymerase III, RNA polymerase II) or other components (for example, an RNA molecule, as occurs in telomerase) will also be understood to be included within the meaning of "protein" as used herein. Similarly, fragments of proteins and polypeptides are also within the scope of the invention and may be referred to herein as "proteins."

[0136] A particular amino acid sequence of a given protein (i.e., the polypeptide's "primary structure," when written from the amino-terminus to carboxy-terminus) is determined by the nucleotide sequence of the coding portion of a mRNA, which is in turn specified by genetic information, typically genomic DNA (including organelle DNA, e.g., mitochondrial or chloroplast DNA). Thus, determining the sequence of a gene assists in predicting the primary sequence of a corresponding polypeptide and more particular the role or activity of the polypeptide or proteins encoded by that gene or polynucleotide sequence.

[0137] The term "isolated" or "purified" when referring to a nucleic acid sequence or a polypeptide sequence, respectively, means altered "by the hand of man" from its natural state; i.e., if it occurs in nature, it has been changed or removed from its original environment, or both. For example, a naturally occurring polynucleotide or a polypeptide naturally present in a living animal, a biological sample or an environmental sample in its natural state is not "isolated" or "purified", but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated" or "purified", as the term is employed herein. Such polynucleotides, when introduced into host cells in culture or in whole organisms, still would be isolated, as the term is used herein, because they would not be in their naturally occurring form or environment. Similarly, the polynucleotides and polypeptides may occur in a composition, such as a media formulation (solutions for introduction of polynucleotides or polypeptides, for example, into cells or compositions or solutions for chemical or enzymatic reactions).

[0138] "Polynucleotide" or "nucleic acid sequence" refers to a polymeric form of nucleotides. In some instances a polynucleotide refers to a sequence that is not immediately contiguous with either of the coding sequences with which it is immediately contiguous (one on the 5' end and one on the 3' end) in the naturally occurring genome of the organism

Where the DNA is genomic DNA, the DNA can be sheared using, for example, a 25 gauge needle.

[0146] The nucleic acids can be cloned into an appropriate vector. The vector used will depend upon whether the DNA is to be expressed, amplified, sequenced or manipulated in any number of ways known in the art (see, for example, U.S. Pat. No. 6,022,716 which discloses high throughput sequencing vectors). Cloning techniques are known in the art or can be developed by one skilled in the art, without undue experimentation. The choice of a vector will also depend on the size of the polynucleotide sequence and the host cell to be employed in the methods of the invention. Thus, the vector used in the invention may be plasmids, phages, cosmids, phagemids, viruses (e.g., retroviruses, parainfluenzavirus, herpesviruses, reoviruses, paramyxoviruses, and the like), or selected portions thereof (e.g., coat protein, spike glycoprotein, capsid protein). For example, cosmids and phagemids are typically used where the specific nucleic acid sequence to be analyzed or modified is large because these vectors are able to stably propagate large polynucleotides.

[0147] The vector containing the cloned nucleic acid sequence can then be amplified by plating (i.e., clonal amplification) or transfecting a suitable host cell with the vector (e.g. a phage on an *E. coli* host). The cloned nucleic acid sequence is used to prepare a library for screening (e.g., expression screening, PCR screening, hybridization screening or the like) by transforming a suitable organism. Hosts, known in the art are transformed by artificial introduction of the vectors containing the nucleic acid sequence by inoculation under conditions conducive for such transformation. One could transform with double stranded circular or linear nucleic acid or there may also be instances where one would transform with single stranded circular or linear nucleic acid sequences. By transform or transformation is meant a permanent or transient genetic change induced in a cell following incorporation of new DNA (e.g., DNA exogenous to the cell). Where the cell is a mammalian cell, a permanent genetic change is generally achieved by introduction of the DNA into the genome of the cell. A transformed cell or host cell generally refers to a cell (e.g., prokaryotic or eukaryotic) into which (or into an ancestor of which) has been introduced, by means of recombinant DNA techniques, a DNA molecule not normally present in the host organism.

[0148] An exemplary vector for use in the invention contains an f-factor origin replication. The f-factor (or fertility factor) in *E. coli* is a plasmid which effects high frequency transfer of itself during conjugation and less frequent transfer of the bacterial chromosome itself. In a particular aspect cloning vectors referred to as "fosmids" or bacterial artificial chromosome (BAC) vectors are used. These are derived from *E. coli* f-factor which is able to stably integrate large segments of DNA. When integrated with DNA from a mixed uncultured environmental sample, this makes it possible to achieve large genomic fragments in the form of a stable environmental gene library.

[0149] The nucleic acids derived from a mixed population or sample may be inserted into the vector by a variety of procedures. In general, the nucleic acid sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art. Such procedures and others are deemed to be within the scope of those skilled in the art. A typical cloning scenario may have DNA "blunted" with an appropriate nuclease (e.g., Mung Bean Nuclease), methylated with, for example, EcoR I Methylase and ligated to EcoR I linkers GGAATTCC (SEQ ID NO:1). The linkers are then digested with an EcoR I Restriction Endonuclease and the DNA size fractionated (e.g., using a sucrose gradient). The resulting size fractionated DNA is then ligated into a suitable vector for sequencing, screening or expression (e.g. a lambda vector and packaged using an in vitro lambda packaging extract).

[0150] Transformation of a host cell with recombinant DNA may be carried out by conventional techniques as are well known to those skilled in the art. Where the host is prokaryotic, such as *E. coli,* competent cells which are capable of DNA uptake can be prepared from cells harvested after exponential growth phase and subsequently treated by the $CaCl_2$ method by procedures well known in the art. Alternatively, $MgCl_2$ or RbCl can be used. Transformation can also be performed after forming a protoplast of the host cell or by electroporation.

[0151] When the host is a eukaryote, methods of transfection or transformation with DNA include calcium phosphate co-precipitates, conventional mechanical procedures such as microinjection, electroporation, insertion of a plasmid encased in liposomes, or virus vectors, as well as others known in the art, may be used. Eukaryotic cells can also be co-transfected with a second foreign DNA molecule encoding a selectable marker, such as the herpes simplex thymidine kinase gene. Another method is to use a eukaryotic viral vector, such as simian virus 40 (SV40) or bovine papilloma virus, to transiently infect or transform eukaryotic cells and express the protein. (Eukaryotic Viral Vectors, Cold Spring Harbor Laboratory, Gluzman ed., 1982). The eukaryotic cell may be a yeast cell (e.g., *Saccharomyces cerevisiae*), an insect cell (e.g., Drosophila sp.) or may be a mammalian cell, including a human cell.

[0152] Eukaryotic systems, and mammalian expression systems, allow for post-translational modifications of expressed mammalian proteins to occur. Eukaryotic cells which possess the cellular machinery for processing of the primary transcript, glycosylation, phosphorylation, or secretion of the gene product should be used. Such host cell lines may include, but are not limited to, CHO, VERO, BHK, HeLa, COS, MDCK, Jurkat, HEK-293, and WI38.

[0153] In one aspect, once a library of clones is created using any number of methods, including those described above, the clones are resuspended in a liquid media, for example, a nutrient rich broth or other growth media known in the art. Typically the media is a liquid media, which can be readily pipetted. One or more media types containing at least one clone of the library are then introduced either individually or together as a mixture, into capillaries (all or a portion thereof) in a capillary array.

[0154] In another aspect, the library is first biopanned prior to introduction or delivery into a capillary device or other screening techniques. Such biopanning methods enrich the library for sequences or activities of interest. Examples of methods for biopanning or enrichment are described below.

[0155] In one aspect, the library can be screened or sorted to enrich for clones containing a sequence or activity of

lecithin, fatty amines and the like. A mixture of fatty material may be employed such a combination of neutral steroid, a charge amphiphile and a phospholipid. Illustrative examples of phospholipids include lecithin, sphingomyelin and dipalmitoylphos-phatidylcholine. Representative steroids include cholesterol, cholestanol and lanosterol. Representative charged amphiphilic compounds generally contain from 12-30 carbon atoms. Mono- or dialkyl phosphate esters, or alkyl amines; e.g., dicetyl phosphate, stearyl amine, hexadecyl amine, dilauryl phosphate, and the like.

**[0166]** Further, it is possible to combine some or all of the above aspects such that a normalization step is performed prior to generation of the expression library, the expression library is then generated, the expression library so generated is then biopanned, and the biopanned expression library is then screened using a high throughput cell sorting and screening instrument. Thus there are a variety of options, including: (i) generating the library and then screen it; (ii) normalize the target DNA, generate the library and screen it; (iii) normalize, generate the library, biopan and screen; or (iv) generate, biopan and screen the library. The nucleic acids used to generate a library can be obtained, for example, from environmental samples, mixed populations of organisms (e.g., cultured or uncultured), enriched populations thereof, and isolates thereof. In addition, the screening techniques include, for example, hybridization screening, PCR screening, expression screening, and the like.

**[0167]** Gel microdroplet or other related technologies can be used in the present invention to localize, sort as well as amplify signals in the high throughput screening of recombinant libraries. Cell viability during the screening is not an issue or concern since nucleic acid can be recovered from the microdroplet.

**[0168]** The biopanning approach described above can be used to create libraries enriched with clones carrying sequences homologous to a given probe sequence. Using this approach libraries containing clones with inserts of up to 40 kbp can be enriched approximately 1,000 fold after each round of panning. This enables one to reduce the number of clones to be screened after 1 round of biopanning enrichment. This approach can be applied to create libraries enriched for clones carrying sequence of interest related to a bioactivity of interest for example polyketide sequences.

**[0169]** Hybridization screening using high-density filters or biopanning has proven an efficient approach to detect homologues of pathways containing conserved genes. To discover novel bioactive molecules that may have no known counterparts, however, other approaches are necessary. One approach is to employ the transaminase screening method described above. Another approach of the present invention is to screen in *E. coli* for the expression of small molecule ring structures or "backbones". Because the genes encoding these polycyclic structures can often be expressed in *E. coil* the small molecule backbone can be manufactured albeit in an inactive form. Bioactivity is conferred upon transferring the molecule or pathway to an appropriate host that expresses the requisite glycosylation and methylation genes that can modify or "decorate" the structure to its active form. Thus, inactive ring compounds, recombinantly expressed in *E. coli* are detected to identify clones, which are then shuttled to a metabolically rich host, such as Streptomyces, for subsequent production of the bioactive molecule. The use of high throughput robotic systems allows the screening of hundreds of thousands of clones in multiplexed arrays in microtiter dishes.

**[0170]** One approach to detect and enrich for clones carrying these structures is to use the capillary screening methods or FACS screening, a procedure described and exemplified in U.S. Ser. No. 08/876,276, filed Jun. 16, 1997. Polycyclic ring compounds typically have characteristic fluorescent spectra when excited by ultraviolet light. Thus, clones expressing these structures can be distinguished from background using a sufficiently sensitive detection method. For example, high throughput FACS screening can be utilized to screen for small molecule backbones in *E. coli* libraries. Commercially available FACS machines are capable of screening up to 100,000 clones per second for UV active molecules. These clones can be sorted for further FACS screening or the resident plasmids can be extracted and shuttled to Streptomyces for activity screening.

**[0171]** In an alternate screening approach, after shuttling to Streptomyces hosts, organic extracts from candidate clones can be tested for bioactivity by susceptibility screening against test organisms such as Staphylococcus aureus, *E. coli*, or *Saccharomyces cervisiae*. FACS screening can be used in this approach by co-encapsulating clones with the test organism.

**[0172]** An alternative to the above-mentioned screening methods provided by the present invention is an approach termed "mixed extract" screening. The "mixed extract" screening approach takes advantage of the fact that the accessory genes needed to confer activity upon the polycyclic backbones are expressed in metabolically rich hosts, such as Streptomyces, and that the enzymes can be extracted and combined with the backbones extracted from *E. coli* clones to produce the bioactive compound in vitro. Enzyme extract preparations from metabolically rich hosts, such as Streptomyces strains, at various growth stages are combined with pools of organic extracts from *E. coli* libraries and then evaluated for bioactivity.

**[0173]** Another approach to detect activity in the *E. coli* clones is to screen for genes that can convert bioactive compounds to different forms. For example, a recombinant enzyme was recently discovered that can convert the low value daunomycin to the higher value doxorubicin. Similar enzyme pathways are being sought to convert penicillins to cephalosporins.

**[0174]** Capillary screening, for example, can also be used to detect expression of UV fluorescent molecules in metabolically rich hosts, such as Streptomyces. Recombinant oxytetracylin retains its diagnostic red fluorescence when produced heterologously in *S. lividans* TK24. Pathway clones, which can be identified by the methods and systems of the invention, can thus be screened for polycyclic molecules in a high throughput fashion.

**[0175]** Recombinant bioactive compounds can also be screened in vivo using "two-hybrid" systems, which can detect enhancers and inhibitors of protein-protein or other interactions such as those between transcription factors and their activators, or receptors and their cognate targets. In this aspect, both a small molecule pathway and a GFP reporter construct are co-expressed. Clones altered in GFP expression can then be identified and the clone isolated for characterization.

[0191] It is also contemplated that populations of double-stranded randomly broken polynucleotides may be mixed or combined at this step with the polynucleotides from the sexual PCR process and optionally subjected to one or more additional sexual PCR cycles.

[0192] Where insertion of mutations into the template polynucleotide is desired, single-stranded or double-stranded polynucleotides having a region of identity to the template polynucleotide and a region of heterology to the template polynucleotide may be added in a 20 fold excess by weight as compared to the total nucleic acid, or the single-stranded polynucleotides may be added in a 10 fold excess by weight as compared to the total nucleic acid.

[0193] Where a mixture of different but related template polynucleotides is desired, populations of polynucleotides from each of the templates may be combined at a ratio of less than about 1:100, or the ratio is less than about 1:40. For example, a backcross of the wild-type polynucleotide with a population of mutated polynucleotide may be desired to eliminate neutral mutations (e.g., mutations yielding an insubstantial alteration in the phenotypic property being selected for). In such an example, the ratio of randomly provided wild-type polynucleotides which may be added to the randomly provided sexual PCR cycle hybrid polynucleotides is approximately 1:1 to about 100:1, and more preferably from 1:1 to 40:1.

[0194] The mixed population of random polynucleotides are denatured to form single-stranded polynucleotides and then re-annealed. Only those single-stranded polynucleotides having regions of homology with other single-stranded polynucleotides will re-anneal.

[0195] Heating may be used to denature the random polynucleotides. One skilled in the art could determine the conditions necessary to completely denature the double-stranded nucleic acid. The temperature can be from 80° C. to 100° C., or the temperature is from 90° C. to 96° C. Other methods, which may be used to denature the polynucleotides include pressure and pH.

[0196] The polynucleotides may be re-annealed by cooling. In one aspect, the temperature is from 20° C. to 75° C., or the temperature is from 40° C. to 65° C. If a high frequency of crossovers is needed based on an average of only 4 consecutive bases of homology, recombination can be forced by using a low annealing temperature, although the process becomes more difficult. The degree of renaturation, which occurs, will depend on the degree of homology between the populations of single-stranded polynucleotides.

[0197] Renaturation can be accelerated by the addition of polyethylene glycol ("PEG") or salt. The salt concentration is preferably from 0 mM to 200 mM, more preferably the salt concentration is from 10 mM to 100 mm. The salt may be KCl or NaCl. The concentration of PEG is preferably from 0% to 20%, more preferably from 5% to 10%.

[0198] The annealed polynucleotides are next incubated in the presence of a nucleic acid polymerase and dNTP's (i.e. dATP, dCTP, DGTP and dTTP). The nucleic acid polymerase may be the Klenow fragment, the Taq polymerase or any other DNA polymerase known in the art.

[0199] The approach to be used for the assembly depends on the minimum degree of homology that should still yield crossovers. If the areas of identity are large, Taq polymerase can be used with an annealing temperature of between 45-65° C. If the areas of identity are small, Klenow polymerase can be used with an annealing temperature of between 20-30° C. One skilled in the art could vary the temperature of annealing to increase the number of crossovers achieved.

[0200] The polymerase may be added to the random polynucleotides prior to annealing, simultaneously with annealing or after annealing.

[0201] The cycle of denaturation, renaturation and incubation in the presence of polymerase is referred to herein as shuffling or reassembly of the nucleic acid. This cycle is repeated for a desired number of times. The cycle can repeated from 2 to 50 times, or more, or the sequence is repeated from 10 to 40 times.

[0202] The resulting nucleic acid is a larger double-stranded polynucleotide of from about 50 bp to about 100 kb, or more, or, the larger polynucleotide is from 500 bp to 50 kb.

[0203] This larger polynucleotides may contain a number of copies of a polynucleotide having the same size as the template polynucleotide in tandem. This concatemeric polynucleotide is then denatured into single copies of the template polynucleotide. The result will be a population of polynucleotides of approximately the same size as the template polynucleotide. The population will be a mixed population where single or double-stranded polynucleotides having an area of identity and an area of heterology have been added to the template polynucleotide prior to shuffling. These polynucleotides are then cloned into the appropriate vector and the ligation mixture used to transform bacteria.

[0204] It is contemplated that the single polynucleotides may be obtained from the larger concatemeric polynucleotide by amplification of the single polynucleotide prior to cloning by a variety of methods including PCR (U.S. Pat. Nos. 4,683,195 and 4,683,202), rather than by digestion of the concatamer.

[0205] The vector used for cloning is not critical provided that it will accept a polynucleotide of the desired size. If expression of the particular polynucleotide is desired, the cloning vehicle should further comprise transcription and translation signals next to the site of insertion of the polynucleotide to allow expression of the polynucleotide in the host cell.

[0206] The resulting bacterial population will include a number of recombinant polynucleotides having random mutations. This mixed population may be tested to identify the desired recombinant polynucleotides. The method of selection will depend on the polynucleotide desired.

[0207] For example, if a polynucleotide, identified by the methods of described herein, encodes a protein with a first binding affinity, subsequent mutated (e.g., shuffled) sequences having an increased binding efficiency to a ligand may be desired. In such a case the proteins expressed by each of the portions of the polynucleotides in the population or library may be tested for their ability to bind to the ligand by methods known in the art (i.e. panning, affinity chromatography). If a polynucleotide, which encodes for a protein with increased drug resistance is desired, the proteins

be used to create scaffold-like proteins with various combinations of mutated sequences for binding.

[0220] The equivalents of some standard genetic matings may also be performed by shuffling in vitro. For example, a "molecular backcross" can be performed by repeatedly mixing the hybrid's nucleic acid with the wild-type nucleic acid while selecting for the mutations of interest. As in traditional breeding, this approach can be used to combine phenotypes from different sources into a background of choice. It is useful, for example, for the removal of neutral mutations that affect unselected characteristics (e.g., immunogenicity). Thus it can be useful to determine which mutations in a protein are involved in the enhanced biological activity and which are not, an advantage which cannot be achieved by error-prone mutagenesis or cassette mutagenesis methods.

[0221] Large, functional genes can be assembled correctly from a mixture of small random polynucleotides. This reaction may be of use for the reassembly of genes from the highly fragmented DNA of fossils. In addition random nucleic acid fragments from fossils may be combined with polynucleotides from similar genes from related species.

[0222] It is also contemplated that the method of the invention can be used for the in vitro amplification of a whole genome from a single cell as is needed for a variety of research and diagnostic applications. DNA amplification by PCR typically includes sequences of about 40 kb. Amplification of a whole genome such as that of *E. coli* (5,000 kb) by PCR would require about 250 primers yielding 125 forty kb polynucleotides. On the other hand, random production of polynucleotides of the genome with sexual PCR cycles, followed by gel purification of small polynucleotides will provide a multitude of possible primers. Use of this mix of random small polynucleotides as primers in a PCR reaction alone or with the whole genome as the template should result in an inverse chain reaction with the theoretical endpoint of a single concatamer containing many copies of the genome.

[0223] A 100-fold amplification in the copy number and an average polynucleotide size of greater than 50 kb may be obtained when only random polynucleotides are used. It is thought that the larger concatamer is generated by overlap of many smaller polynucleotides. The quality of specific PCR products obtained using synthetic primers will be indistinguishable from the product obtained from unamplified DNA. It is expected that this approach will be useful for the mapping of genomes.

[0224] The polynucleotide to be shuffled can be produced as random or non-random polynucleotides, at the discretion of the practitioner. Moreover, the invention provides a method of shuffling that is applicable to a wide range of polynucleotide sizes and types, including the step of generating polynucleotide monomers to be used as building blocks in the reassembly of a larger polynucleotide. For example, the building blocks can be fragments of genes or they can be comprised of entire genes or gene pathways, or any combination thereof.

[0225] In an aspect of in vivo shuffling, a mixed population of a specific nucleic acid sequence is introduced into bacterial or eukaryotic cells under conditions such that at least two different nucleic acid sequences are present in each host cell. The polynucleotides can be introduced into the host cells by a variety of different methods. The host cells can be transformed with the smaller polynucleotides using methods known in the art, for example treatment with calcium chloride. If the polynucleotides are inserted into a phage genome, the host cell can be transfected with the recombinant phage genome having the specific nucleic acid sequences. Alternatively, the nucleic acid sequences can be introduced into the host cell using electroporation, transfection, lipofection, biolistics, conjugation, and the like.

[0226] In general, in this aspect, specific nucleic acid sequences will be present in vectors, which are capable of stably replicating the sequence in the host cell. In addition, it is contemplated that the vectors will encode a marker gene such that host cells having the vector can be selected. This ensures that the mutated specific nucleic acid sequence can be recovered after introduction into the host cell. However, it is contemplated that the entire mixed population of the specific nucleic acid sequences need not be present on a vector sequence. Rather only a sufficient number of sequences need be cloned into vectors to ensure that after introduction of the polynucleotides into the host cells each host cell contains one vector having at least one specific nucleic acid sequence present therein. It is also contemplated that rather than having a subset of the population of the specific nucleic acids sequences cloned into vectors, this subset may be already stably integrated into the host cell.

[0227] It has been found that when two polynucleotides, which have regions of identity are inserted into the host cells homologous recombination occurs between the two polynucleotides. Such recombination between the two mutated specific nucleic acid sequences will result in the production of double or triple hybrids in some situations.

[0228] It has also been found that the frequency of recombination is increased if some of the mutated specific nucleic acid sequences are present on linear nucleic acid molecules. Therefore, in a one aspect, some of the specific nucleic acid sequences are present on linear polynucleotides.

[0229] After transformation, the host cell transformants are placed under selection to identify those host cell transformants which contain mutated specific nucleic acid sequences having the qualities desired. For example, if increased resistance to a particular drug is desired then the transformed host cells may be subjected to increased concentrations of the particular drug and those transformants producing mutated proteins able to confer increased drug resistance will be selected. If the enhanced ability of a particular protein to bind to a receptor is desired, then expression of the protein can be induced from the transformants and the resulting protein assayed in a ligand binding assay by methods known in the art to identify that subset of the mutated population which shows enhanced binding to the ligand. Alternatively, the protein can be expressed in another system to ensure proper processing.

[0230] Once a subset of the first recombined specific nucleic acid sequences (daughter sequences) having the desired characteristics are identified, they are then subject to a second round of recombination. In the second cycle of recombination, the recombined specific nucleic acid sequences may be mixed with the original mutated specific nucleic acid sequences (parent sequences) and the cycle repeated as described above. In this way a set of second recombined specific nucleic acids sequences can be identi-

invention. An exonuclease can be used to liberate nucleotides from one or both ends of a linear double stranded polynucleotide, and from one to all ends of a branched polynucleotide having more than two ends.

[0244] By contrast, a non-enzymatic step may be used to shuffle, assemble, reassemble, recombine, and/or concatenate polynucleotide building blocks that is comprised of subjecting a working sample to denaturing (or "melting") conditions (for example, by changing temperature, pH, and /or salinity conditions) so as to melt a working set of double stranded polynucleotides into single polynucleotide strands. For shuffling, it is desirable that the single polynucleotide strands participate to some extent in annealment with different hybridization partners (i.e. and not merely revert to exclusive re-annealment between what were former partners before the denaturation step). The presence of the former hybridization partners in the reaction vessel, however, does not preclude, and may sometimes even favor, re-annealment of a single stranded polynucleotide with its former partner, to recreate an original double stranded polynucleotide.

[0245] In contrast to this non-enzymatic shuffling step comprised of subjecting double stranded polynucleotide building blocks to denaturation, followed by annealment, the invention further provides an exonuclease-based approach requiring no denaturation - rather, the avoidance of denaturing conditions and the maintenance of double stranded polynucleotide substrates in annealed (i.e. non-denatured) state are necessary conditions for the action of exonucleases (e.g., exonuclease III and red alpha gene product). In further contrast, the generation of single stranded polynucleotide sequences capable of hybridizing to other single stranded polynucleotide sequences is the result of covalent cleavage—and hence sequence destruction—in one of the hybridization partners. For example, an exonuclease III enzyme may be used to enzymatically liberate 3' terminal nucleotides in one hybridization strand (to achieve covalent hydrolysis in that polynucleotide strand); and this favors hybridization of the remaining single strand to a new partner (since its former partner was subjected to covalent cleavage).

[0246] It is particularly appreciated that enzymes can be discovered, optimized (e.g., engineered by directed evolution), or both discovered and optimized specifically for the instantly disclosed approach that have more optimal rates and/or more highly specific activities &/or greater lack of unwanted activities. In fact it is expected that the invention may encourage the discovery and/or development of such designer enzymes.

[0247] Furthermore, it is appreciated that one can protect the end of a double stranded polynucleotide or render it susceptible to a desired enzymatic action of an exonuclease as necessary. For example, a double stranded polynucleotide end having a 3' overhang is not susceptible to the exonuclease action of exonuclease III. However, it may be rendered susceptible to the exonuclease action of exonuclease III by a variety of means; for example, it may be blunted by treatment with a polymerase, cleaved to provide a blunt end or a 5' overhang, joined (ligated or hybridized) to another double stranded polynucleotide to provide a blunt end or a 5' overhang, hybridized to a single stranded polynucleotide to provide a blunt end or a 5' overhang, or modified by any of a variety of means).

[0248] According to one aspect, an exonuclease may be allowed to act on one or on both ends of a linear double stranded polynucleotide and proceed to completion, to near completion, or to partial completion. When the exonuclease action is allowed to go to completion, the result will be that the length of each 5' overhang will be extend far towards the middle region of the polynucleotide in the direction of what might be considered a "rendezvous point" (which may be somewhere near the polynucleotide midpoint). Ultimately, this results in the production of single stranded polynucleotides (that can become dissociated) that are each about half the length of the original double stranded polynucleotide.

[0249] Thus, the exonuclease-mediated approach is useful for shuffling, assembling and/or reassembling, recombining, and concatenating polynucleotide building blocks. The polynucleotide building blocks can be up to ten bases long or tens of bases long or hundreds of bases long or thousands of bases long or tens of thousands of bases long or hundreds of thousands of bases long or millions of bases long or even longer.

[0250] Subjecting a double stranded polynucleotide to fragmentation may be used to generate substrates for an exonuclease. Fragmentation may be achieved by mechanical means (e.g., shearing, sonication, and the like), by enzymatic means (e.g., using restriction enzymes), and by any combination thereof. Fragments of a larger polynucleotide may also be generated by polymerase-mediated synthesis.

[0251] Additional examples of enzymes with exonuclease activity include red-alpha and venom phosphodiesterases. Red alpha (red gene product (also referred to as lambda exonuclease) is of bacteriophage origin. Red alpha gene product acts processively from 5'-phosphorylated termini to liberate mononucleotides from duplex DNA (Takahashi & Kobayashi, 1990). Venom phosphodiesterases (Laskowski, 1980) is capable of rapidly opening supercoiled DNA.

[0252] In one aspect, the present invention provides a non-stochastic method termed synthetic ligation reassembly (SLR), that is somewhat related to stochastic shuffling, save that the nucleic acid building blocks are not shuffled or concatenated or chimerized randomly, but rather are assembled non-stochastically.

[0253] The SLR method does not depend on the presence of a high level of homology between polynucleotides to be shuffled. The invention can be used to non-stochastically generate libraries (or sets) of progeny molecules comprised of over $10^{100}$ different chimeras. Conceivably, SLR can even be used to generate libraries comprised of over $10^{1000}$ different progeny chimeras.

[0254] Thus, in one aspect, the invention provides a non-stochastic method of producing a set of finalized chimeric nucleic acid molecules having an overall assembly order that is chosen by design, which method is comprised of the steps of generating, by design, a plurality of specific nucleic acid building blocks having serviceable mutually compatible ligatable ends, and assembling these nucleic acid building blocks, such that a designed overall assembly order is achieved.

[0255] The mutually compatible ligatable ends of the nucleic acid building blocks to be assembled are considered to be "serviceable" for this type of ordered assembly if they enable the building blocks to be coupled in predetermined

functional in one or more host cells for gene splicing much in the way that naturally-occurring introns serve functionally in gene splicing. The invention provides a process of producing man-made intron-containing polynucleotides to be introduced into host organisms for recombination and/or splicing.

[0267] A man-made gene produced using the invention can also serve as a substrate for recombination with another nucleic acid. Likewise, a man-made gene pathway produced using the invention can also serve as a substrate for recombination with another nucleic acid. In one instance, the recombination is facilitated by, or occurs at, areas of homology between the man-made intron-containing gene and a nucleic acid with serves as a recombination partner. The recombination partner may also be a nucleic acid generated by the invention, including a man-made gene or a man-made gene pathway. Recombination may be facilitated by or may occur at areas of homology that exist at the one (or more) artificially introduced intron(s) in the man-made gene.

[0268] The synthetic ligation reassembly method of the invention utilizes a plurality of nucleic acid building blocks, each of which can have two ligatable ends. The two ligatable ends on each nucleic acid building block may be two blunt ends (i.e. each having an overhang of zero nucleotides), or, one blunt end and one overhang, or, two overhangs.

[0269] An overhang for this purpose may be a 3' overhang or a 5' overhang. Thus, a nucleic acid building block may have a 3' overhang or alternatively a 5' overhang or alternatively two 3' overhangs or alternatively two 5' overhangs. The overall order in which the nucleic acid building blocks are assembled to form a finalized chimeric nucleic acid molecule is determined by purposeful experimental design and is not random.

[0270] A nucleic acid building block can be generated by chemical synthesis of two single-stranded nucleic acids (also referred to as single-stranded oligos) and contacting them so as to allow them to anneal to form a double-stranded nucleic acid building block.

[0271] A double-stranded nucleic acid building block can be of variable size. The sizes of these building blocks can be small or large. Alternative sizes for building block range from 1 base pair (not including any overhangs) to 100,000 base pairs (not including any overhangs). Other size ranges are also provided, which have lower limits of from 1 bp to 10,000 bp (including every integer value in between), and upper limits of from 2 bp to 100,000 bp (including every integer value in between).

[0272] Many methods exist by which a double-stranded nucleic acid building block can be generated that is serviceable for the invention; and these are known in the art and can be readily performed by the skilled artisan.

[0273] According to one aspect, a double-stranded nucleic acid building block is generated by first generating two single stranded nucleic acids and allowing them to anneal to form a double-stranded nucleic acid building block. The two strands of a double-stranded nucleic acid building block may be complementary at every nucleotide apart from any that form an overhang; thus containing no mismatches, apart from any overhang(s). According to another aspect, the two strands of a double-stranded nucleic acid building block are complementary at fewer than every nucleotide apart from any that form an overhang. Thus, according to this aspect, a double-stranded nucleic acid building block can be used to introduce codon degeneracy. The codon degeneracy can be introduced using the site-saturation mutagenesis described herein, using one or more N,N,G/T cassettes or alternatively using one or more N,N,N cassettes.

[0274] The in vivo recombination method of the invention can be performed blindly on a pool of unknown hybrids or alleles of a specific polynucleotide or sequence. However, it is not necessary to know the actual DNA or RNA sequence of the specific polynucleotide.

[0275] The approach of using recombination within a mixed population of genes can be useful for the generation of any useful proteins, for example, interleukin I, antibodies, tPA and growth hormone. This approach may be used to generate proteins having altered specificity or activity. The approach may also be useful for the generation of hybrid nucleic acid sequences, for example, promoter regions, introns, exons, enhancer sequences, 31 untranslated regions or 51 untranslated regions of genes. Thus this approach may be used to generate genes having increased rates of expression. This approach may also be useful in the study of repetitive DNA sequences. Finally, this approach may be useful to mutate ribozymes or aptamers.

[0276] The invention provides a method for selecting a subset of polynucleotides from a starting set of polynucleotides, which method is based on the ability to discriminate one or more selectable features (or selection markers) present anywhere in a working polynucleotide, so as to allow one to perform selection for (positive selection) and/or against (negative selection) each selectable polynucleotide. In a one aspect, a method is provided termed end-selection, which method is based on the use of a selection marker located in part or entirely in a terminal region of a selectable polynucleotide, and such a selection marker may be termed an "end-selection marker".

[0277] End-selection may be based on detection of naturally occurring sequences or on detection of sequences introduced experimentally (including by any mutagenesis procedure mentioned herein and not mentioned herein) or on both, even within the same polynucleotide. An end-selection marker can be a structural selection marker or a functional selection marker or both a structural and a functional selection marker. An end-selection marker may be comprised of a polynucleotide sequence or of a polypeptide sequence or of any chemical structure or of any biological or biochemical tag, including markers that can be selected using methods based on the detection of radioactivity, of enzymatic activity, of fluorescence, of any optical feature, of a magnetic property (e.g., using magnetic beads), of immunoreactivity, and of hybridization.

[0278] End-selection may be applied in combination with any method for performing mutagenesis. Such mutagenesis methods include, but are not limited to, methods described herein (supra and infra). Such methods include, by way of non-limiting exemplification, any method that may be referred herein or by others in the art by any of the following terms: "saturation mutagenesis", "shuffling", "recombination", "re-assembly", "error-prone PCR", "assembly PCR", "sexual PCR", "crossover PCR", "oligonucleotide primer-directed mutagenesis", "recursive (and/or exponential) ensemble mutagenesis (see Arkin and Youvan, 1992)", "cas-

identification of peptide structures, including the primary amino acid sequences, of peptides or peptidomimetics that interact with biological macromolecules. One method of identifying peptides that possess a desired structure or functional property, such as binding to a predetermined biological macromolecule (e.g., a receptor), involves the screening of a large library or peptides for individual library members which possess the desired structure or functional property conferred by the amino acid sequence of the peptide.

[0289] In addition to direct chemical synthesis methods for generating peptide libraries, several recombinant DNA methods also have been reported. One type involves the display of a peptide sequence, antibody, or other protein on the surface of a bacteriophage particle or cell. Generally, in these methods each bacteriophage particle or cell serves as an individual library member displaying a single species of displayed peptide in addition to the natural bacteriophage or cell protein sequences. Each bacteriophage or cell contains the nucleotide sequence information encoding the particular displayed peptide sequence; thus, the displayed peptide sequence can be ascertained by nucleotide sequence determination of an isolated library member.

[0290] A well-known peptide display method involves the presentation of a peptide sequence on the surface of a filamentous bacteriophage, typically as a fusion with a bacteriophage coat protein. The bacteriophage library can be incubated with an immobilized, predetermined macromolecule or small molecule (e.g., a receptor) so that bacteriophage particles which present a peptide sequence that binds to the immobilized macromolecule can be differentially partitioned from those that do not present peptide sequences that bind to the predetermined macromolecule. The bacteriophage particles (i.e., library members), which are bound to the immobilized macromolecule are then recovered and replicated to amplify the selected bacteriophage sub-population for a subsequent round of affinity enrichment and phage replication. After several rounds of affinity enrichment and phage replication, the bacteriophage library members that are thus selected are isolated and the nucleotide sequence encoding the displayed peptide sequence is determined, thereby identifying the sequence(s) of peptides that bind to the predetermined macromolecule (e.g., receptor). Such methods are further described in PCT patent publications WO 91/17271, WO 91/18980, WO 91/19818 and WO 93/08278.

[0291] The present invention also provides random, pseudorandom, and defined sequence framework peptide libraries and methods for generating and screening those libraries to identify useful compounds (e.g., peptides, including single-chain antibodies) that bind to receptor molecules or epitopes of interest or gene products that modify peptides or RNA in a desired fashion. The random, pseudorandom, and defined sequence framework peptides are produced from libraries of peptide library members that comprise displayed peptides or displayed single-chain antibodies attached to a polynucleotide template from which the displayed peptide was synthesized. The mode of attachment may vary according to the specific aspect of the invention selected, and can include encapsulation in a phage particle or incorporation in a cell.

[0292] One advantage of the present invention is that no prior information regarding an expected ligand structure is required to isolate peptide ligands or antibodies of interest. The peptide identified can have biological activity, which is meant to include at least specific binding affinity for a selected receptor molecule and, in some instances, will further include the ability to block the binding of other compounds, to stimulate or inhibit metabolic pathways, to act as a signal or messenger, to stimulate or inhibit cellular activity, and the like.

[0293] The invention also provides a method for shuffling a pool of polynucleotide sequences identified by the methods of the invention and selected by affinity screening a library of polysomes displaying nascent peptides (including single-chain antibodies) for library members which bind to a predetermined receptor (e.g., a mammalian proteinaceous receptor such as, for example, a peptidergic hormone receptor, a cell surface receptor, an intracellular protein which binds to other protein(s) to form intracellular protein complexes such as hetero-dimers and the like) or epitope (e.g., an immobilized protein, glycoprotein, oligosaccharide, and the like).

[0294] Polynucleotide sequences selected in a first selection round (typically by affinity selection for binding to a receptor (e.g., a ligand)) by any of these methods are pooled and the pool(s) is/are shuffled by in vitro and/or in vivo recombination to produce a shuffled pool comprising a population of recombined selected polynucleotide sequences. The recombined selected polynucleotide sequences are subjected to at least one subsequent selection round. The polynucleotide sequences selected in the subsequent selection round(s) can be used directly, sequenced, and/or subjected to one or more additional rounds of shuffling and subsequent selection. Selected sequences can also be back-crossed with polynucleotide sequences encoding neutral sequences (i.e., having insubstantial functional effect on binding), such as for example by back-crossing with a wild-type or naturally-occurring sequence substantially identical to a selected sequence to produce native-like functional peptides, which may be less immunogenic. Generally, during back-crossing subsequent selection is applied to retain the property of binding to the predetermined receptor (ligand).

[0295] Prior to or concomitant with the shuffling of selected sequences, the sequences can be mutagenized. In one aspect, selected library members are cloned in a prokaryotic vector (e.g., plasmid, phagemid, or bacteriophage) wherein a collection of individual colonies (or plaques) representing discrete library members are produced. Individual selected library members can then be manipulated (e.g., by site-directed mutagenesis, cassette mutagenesis, chemical mutagenesis, PCR mutagenesis, and the like) to generate a collection of library members representing a kernal of sequence diversity based on the sequence of the selected library member. The sequence of an individual selected library member or pool can be manipulated to incorporate random mutation, pseudorandom mutation, defined kernal mutation (i.e., comprising variant and invariant residue positions and/or comprising variant residue positions which can comprise a residue selected from a defined subset of amino acid residues), codon-based mutation, and the like, either segmentally or over the entire length of the individual selected library member sequence. The

subsequent washing. The temperature, pH, ionic strength, divalent cations concentration, and the volume and duration of the washing will select for nascent peptide/DNA complexes within particular ranges of affinity for the immobilized macromolecule. Selection based on slow dissociation rate, which is usually predictive of high affinity, is often the most practical route. This may be done either by continued incubation in the presence of a saturating amount of free predetermined macromolecule, or by increasing the volume, number, and length of the washes. In each case, the rebinding of dissociated nascent peptide/DNA or peptide/RNA complex is prevented, and with increasing time, nascent peptide/DNA or peptide/RNA complexes of higher and higher affinity are recovered.

[0305] Additional modifications of the binding and washing procedures may be applied to find peptides with special characteristics. The affinities of some peptides are dependent on ionic strength or cation concentration. This is a useful characteristic for peptides that will be used in affinity purification of various proteins when gentle conditions for removing the protein from the peptides are required.

[0306] One variation involves the use of multiple binding targets (multiple epitope species, multiple receptor species), such that a scfv library can be simultaneously screened for a multiplicity of scfv which have different binding specificities. Given that the size of a scfv library often limits the diversity of potential scfv sequences, it is typically desirable to us scfv libraries of as large a size as possible. The time and economic considerations of generating a number of very large polysome scFv-display libraries can become prohibitive. To avoid this substantial problem, multiple predetermined epitope species (receptor species) can be concomitantly screened in a single library, or sequential screening against a number of epitope species can be used. In one variation, multiple target epitope species, each encoded on a separate bead (or subset of beads), can be mixed and incubated with a polysome-display scfv library under suitable binding conditions. The collection of beads, comprising multiple epitope species, can then be used to isolate, by affinity selection, scfv library members. Generally, subsequent affinity screening rounds can include the same mixture of beads, subsets thereof, or beads containing only one or two individual epitope species. This approach affords efficient screening, and is compatible with laboratory automation, batch processing, and high throughput screening methods.

[0307] A variety of techniques can be used in the present invention to diversify a peptide library or single-chain antibody library, or to diversify, prior to or concomitant with shuffling, around variable segment peptides found in early rounds of panning to have sufficient binding activity to the predetermined macromolecule or epitope. In one approach, the positive selected peptide/polynucleotide complexes (those identified in an early round of affinity enrichment) are sequenced to determine the identity of the active peptides. Oligonucleotides are then synthesized based on these active peptide sequences, employing a low level of all bases incorporated at each step to produce slight variations of the primary oligonucleotide sequences. This mixture of (slightly) degenerate oligonucleotides is then cloned into the variable segment sequences at the appropriate locations. This method produces systematic, controlled variations of the starting peptide sequences, which can then be shuffled. It requires, however, that individual positive nascent peptide/polynucleotide complexes be sequenced before mutagenesis, and thus is useful for expanding the diversity of small numbers of recovered complexes and selecting variants having higher binding affinity and/or higher binding specificity. In a variation, mutagenic PCR amplification of positive selected peptide/polynucleotide complexes (especially of the variable region sequences, the amplification products of which are shuffled in vitro and/or in vivo and one or more additional rounds of screening is done prior to sequencing. The same general approach can be employed with single-chain antibodies in order to expand the diversity and enhance the binding affinity/specificity, typically by diversifying CDRs or adjacent framework regions prior to or concomitant with shuffling. If desired, shuffling reactions can be spiked with mutagenic oligonucleotides capable of in vitro recombination with the selected library members can be included. Thus, mixtures of synthetic oligonucleotides and PCR produced polynucleotides (synthesized by error-prone or high-fidelity methods) can be added to the in vitro shuffling mix and be incorporated into resulting shuffled library members (shufflants).

[0308] The invention of shuffling enables the generation of a vast library of CDR-variant single-chain antibodies. One way to generate such antibodies is to insert synthetic CDRs into the single-chain antibody and/or CDR randomization prior to or concomitant with shuffling. The sequences of the synthetic CDR cassettes are selected by referring to known sequence data of human CDR and are selected in the discretion of the practitioner according to the following guidelines: synthetic CDRs will have at least 40 percent positional sequence identity to known CDR sequences, and preferably will have at least 50 to 70 percent positional sequence identity to known CDR sequences. For example, a collection of synthetic CDR sequences can be generated by synthesizing a collection of oligonucleotide sequences on the basis of naturally-occurring human CDR sequences listed in Kabat (Kabat et al., 1991); the pool (s) of synthetic CDR sequences are calculated to encode CDR peptide sequences having at least 40 percent sequence identity to at least one known naturally-occurring human CDR sequence. Alternatively, a collection of naturally-occurring CDR sequences may be compared to generate consensus sequences so that amino acids used at a residue position frequently (i.e., in at least 5 percent of known CDR sequences) are incorporated into the synthetic CDRs at the corresponding position(s). Typically, several (e.g., 3 to about 50) known CDR sequences are compared and observed natural sequence variations between the known CDRs are tabulated, and a collection of oligonucleotides encoding CDR peptide sequences encompassing all or most permutations of the observed natural sequence variations is synthesized. For example but not for limitation, if a collection of human VH CDR sequences have carboxy-terminal amino acids which are either Tyr, Val, Phe, or Asp, then the pool(s) of synthetic CDR oligonucleotide sequences are designed to allow the carboxy-terminal CDR residue to be any of these amino acids. In some aspects, residues other than those which naturally-occur at a residue position in the collection of CDR sequences are incorporated: conservative amino acid substitutions are frequently incorporated and up to 5 residue positions may be varied to incorporate non-conservative amino acid substitutions as compared to known naturally-occurring CDR sequences. Such CDR sequences

1993; Guarente, 1993) can be used to identify interacting protein sequences (i.e., protein sequences which heterodimerize or form higher order heteromultimers). Sequences selected by a two-hybrid system can be pooled and shuffled and introduced into a two-hybrid system for one or more subsequent rounds of screening to identify polypeptide sequences which bind to the hybrid containing the predetermined binding sequence. The sequences thus identified can be compared to identify consensus sequence(s) and consensus sequence kernals.

[0315] One microgram samples of template DNA are obtained and treated with U.V. light to cause the formation of dimers, including TT dimers, particularly purine dimers. U.V. exposure is limited so that only a few photoproducts are generated per gene on the template DNA sample. Multiple samples are treated with U.V. light for varying periods of time to obtain template DNA samples with varying numbers of dimers from U.V. exposure.

[0316] A random priming kit which utilizes a non-proofreading polymerase (for example, Prime-It II Random Primer Labeling kit by Stratagene Cloning Systems) is utilized to generate different size polynucleotides by priming at random sites on templates which are prepared by U.V. light (as described above) and extending along the templates. The priming protocols such as described in the Prime-It II Random Primer Labeling kit may be utilized to extend the primers. The dimers formed by U.V. exposure serve as a roadblock for the extension by the non-proofreading polymerase. Thus, a pool of random size polynucleotides is present after extension with the random primers is finished.

[0317] The invention is further directed to a method for generating a selected mutant polynucleotide sequence (or a population of selected polynucleotide sequences) typically in the form of amplified and/or cloned polynucleotides, whereby the selected polynucleotide sequences(s) possess at least one desired phenotypic characteristic (e.g., encodes a polypeptide, promotes transcription of linked polynucleotides, binds a protein, and the like) which can be selected for. One method for identifying hybrid polypeptides that possess a desired structure or functional property, such as binding to a predetermined biological macromolecule (e.g., a receptor), involves the screening of a large library of polypeptides for individual library members which possess the desired structure or functional property conferred by the amino acid sequence of the polypeptide.

[0318] In one aspect, the present invention provides a method for generating libraries of displayed polypeptides or displayed antibodies suitable for affinity interaction screening or phenotypic screening. The method comprises (1) obtaining a first plurality of selected library members comprising a displayed polypeptide or displayed antibody and an associated polynucleotide encoding said displayed polypeptide or displayed antibody, and obtaining said associated polynucleotides or copies thereof wherein said associated polynucleotides comprise a region of substantially identical sequences, optimally introducing mutations into said polynucleotides or copies, (2) pooling the polynucleotides or copies, (3) producing smaller or shorter polynucleotides by interrupting a random or particularized priming and synthesis process or an amplification process, and (4) performing amplification, preferably PCR amplification, and optionally mutagenesis to homologously recombine the newly synthesized polynucleotides.

[0319] It is an object of the invention to provide a process for producing hybrid polynucleotides, which express a useful hybrid polypeptide by a series of steps comprising:

[0320] (a) producing polynucleotides by interrupting a polynucleotide amplification or synthesis process with a means for blocking or interrupting the amplification or synthesis process and thus providing a plurality of smaller or shorter polynucleotides due to the replication of the polynucleotide being in various stages of completion;

[0321] (b) adding to the resultant population of single- or double-stranded polynucleotides one or more single- or double-stranded oligonucleotides, wherein said added oligonucleotides comprise an area of identity in an area of heterology to one or more of the single- or double-stranded polynucleotides of the population;

[0322] (c) denaturing the resulting single- or double-stranded oligonucleotides to produce a mixture of single-stranded polynucleotides, optionally separating the shorter or smaller polynucleotides into pools of polynucleotides having various lengths and further optionally subjecting said polynucleotides to a PCR procedure to amplify one or more oligonucleotides comprised by at least one of said polynucleotide pools;

[0323] (d) incubating a plurality of said polynucleotides or at least one pool of said polynucleotides with a polymerase under conditions which result in annealing of said single-stranded polynucleotides at regions of identity between the single-stranded polynucleotides and thus forming of a mutagenized double-stranded polynucleotide chain;

[0324] (e) optionally repeating steps (c) and (d);

[0325] (f) expressing at least one hybrid polypeptide from said polynucleotide chain, or chains; and

[0326] (g) screening said at least one hybrid polypeptide for a useful activity.

[0327] In a one aspect of the invention, the means for blocking or interrupting the amplification or synthesis process is by utilization of UV light, DNA adducts, DNA binding proteins.

[0328] In one aspect of the invention, the DNA adducts, or polynucleotides comprising the DNA adducts, are removed from the polynucleotides or polynucleotide pool, such as by a process including heating the solution comprising the DNA fragments prior to further processing.

[0329] In another aspect, clones, which are identified as having a biomolecule or bioactivity of interest, may also be sequenced to identify the DNA sequence encoding a polypeptide (e.g., an enzyme) or the polypeptide sequence itself having the specified activity, for example. Thus, in accordance with the present invention it is possible to isolate and identify: (i) DNA encoding a bioactivity of interest (e.g., an enzyme having a specified enzyme activity), (ii) biomolecules (e.g., polynucleotides or enzymes having such activ-

[0338] The BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, Proc. Natl. Acad. Sci. USA 90:5873 (1993)). One measure of similarity provided by BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide sequences would occur by chance. For example, a nucleic acid is considered similar to a references sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.2, more preferably less than about 0.01, and most preferably less than about 0.001.

[0339] Sequence homology means that two polynucleotide sequences are homologous (i.e., on a nucleotide-by-nucleotide basis) over the window of comparison. A percentage of sequence identity or homology is calculated by comparing two optimally aligned sequences over the window of comparison, determining the number of positions at which the identical nucleic acid base (e.g., A, T, C, G, U, or I) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison (i.e., the window size), and multiplying the result by 100 to yield the percentage of sequence homology. This substantial homology denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence having at least 60 percent sequence homology, typically at least 70 percent homology, often 80 to 90 percent sequence homology, and most commonly at least 99 percent sequence homology as compared to a reference sequence of a comparison window of at least 25-50 nucleotides, wherein the percentage of sequence homology is calculated by comparing the reference sequence to the polynucleotide sequence which may include deletions or additions which total 20 percent or less of the reference sequence over the window of comparison.

[0340] Sequences having sufficient homology can be further identified by any annotations contained in the database, including, for example, species and activity information. Accordingly, in a typical environmental sample, a plurality of nucleic acid sequences will be obtained, cloned, sequenced and corresponding homologous sequences from a database identified. This information provides a profile of the polynucleotides present in the sample, including one or more features associated with the polynucleotide including the organism and activity associated with that sequence or any polypeptide encoded by that sequence based on the database information. As used herein "fingerprint" or "profile" refers to the fact that each sample will have associated with it a set of polynucleotides characteristic of the sample and the environment from which it was derived. Such a profile can include the amount and type of sequences present in the sample, as well as information regarding the potential activities encoded by the polynucleotides and the organisms from which polynucleotides were derived. This unique pattern is each sample's profile or fingerprint.

[0341] In some instances it may be desirable to express a particular cloned polynucleotide sequence once its identity or activity is determined or a suggested identity or activity is associated with the polynucleotide. In such instances the desired clone, if not already cloned into an expression vector, is ligated downstream of a regulatory control element (e.g., a promoter or enhancer) and cloned into a suitable host cell. Expression vectors are commercially available along with corresponding host cells for use in the invention.

[0342] As representative examples of expression vectors which may be used there may be mentioned viral particles, baculovirus, phage, plasmids, phagemids, cosmids, fosmids, bacterial artificial chromosomes, viral nucleic acid (e.g., vaccinia, adenovirus, foul pox virus, pseudorabies and derivatives of SV40), P1-based artificial chromosomes, yeast plasmids, yeast artificial chromosomes, and any other vectors specific for specific hosts of interest (such as bacillus, Aspergillus, yeast, and the like). Thus, for example, the DNA may be included in any one of a variety of expression vectors for expressing a polypeptide. Such vectors include chromosomal, nonchromosomal and synthetic DNA sequences. Large numbers of suitable vectors are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example; Bacterial: pQE70, pQE60, pQE-9 (Qiagen), psiX174, pBluescript SK, pBluescript KS, pNH8A, pNH16a, pNH 18A, pNH46A (Stratagene); pTRC99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); Eukaryotic: pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene), pSVK3, pBPV, pMSG, pSVL (Pharmacia). However, any other plasmid or vector may be used as long as they are replicable and viable in the host.

[0343] The nucleic acid sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. Particular named bacterial promoters include lac, lacZ, T3, T7, gpt, lambda PR, PL and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression. Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers.

[0344] In addition, the expression vectors typically contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E. coli*.

[0345] The nucleic acid sequence(s) selected, cloned and sequenced as hereinabove described can additionally be introduced into a suitable host to prepare a library, which is screened for the desired biomolecule or bioactivity. The selected nucleic acid is preferably already in a vector which includes appropriate control sequences whereby a selected nucleic acid encoding a biomolecule or bioactivity may be expressed, for detection of the desired activity. The host cell can be a higher eukaryotic cell, such as a mammalian cell, or a lower eukaryotic cell, such as a yeast cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. The selection of an appropriate host is deemed to be within the scope of those skilled in the art from the teachings herein.

[0346] In some instances it may be desirable to perform an amplification of the nucleic acid sequence present in a sample or a particular clone that has been isolated. In this

plasmid-borne ADH. These strains have been engineered to be compatible with any library for use in selections.

[0361] Strains that grow on ethanol under aerobic conditions can be used. In addition, we have found that acetaldehyde serves as a sole C-source for growth. Thus, any enzymatic reaction generating either ethanol or acetaldehyde can be selected for using these strains. Application of these strains to aldolase discovery has already been discussed, above. Another useful example is for the discovery of hydrolytic haloalkane dehalogenases as shown in FIG. 13 (illustrating a selection strategy for novel dehalogenases).

Enantioselective Screening of Epoxide Hydrolases via Selections

[0362] Selection assays, in which only the clones containing activities of interest grow, are very powerful for new enzyme discovery because they offer exceptionally high throughput. This approach is especially powerful when applied to environmental DNA libraries. The invention provides an approach to discover and evolve epoxide hydrolases using a novel selection strategy. The approach relies on utilizing epoxide substrates that can be converted by epoxide hydrolases to diols that can be utilized by host bacteria as a carbon source. When environmental library cells are grown in minimal media supplemented with this epoxide as the sole carbon source, only those clones harboring active epoxide hydrolases will be able to produce the corresponding diol and to utilize it as a carbon source for growth and proliferation. Over time, these clones will dominate the microbial population, and thus can be readily isolated. If the selected diols are chiral and enantiomerically pure, this selection method may be used for screening epoxide hydrolases with preference toward one enantiomeric form over the other, thus resulting the discovery of epoxide hydrolases with high enantioselectivity.

[0363] For example, two chiral epoxides, glycidol and propylene oxide, can be used as selection substrates for the above described purpose, see FIG. 14. The corresponding vicinal diols, glycerol, propane diol, are known to support the growth of *E. coli* or its mutants as sole carbon sources. When pure enantiomers of glycidol or propylene oxide are used for selections, epoxide hydrolases with selectivity for these enantiomers may be discovered.

[0364] Appropriate hosts need to be developed for the selection experiments. For example, an *E. coli* mutant with mutations in the glycerol metabolic pathway genes glpK and glpR may be required to facilitate glycerol utilization (see, e.g., Maloy, S. R.; Nunn, W. D. J. Bacteriol. 1982, 149, 173-180). See FIG. 15, illustrating the glycerol metabolic pathway.

[0365] Also, an *E. coli* mutant with a constitutively expressed oxidoreductase (fucO) that can convert propane diol to lactaldehyde is required for propylene oxide to be used as carbon source for selection (see, e.g., Hacking, A. J.; Lin, E. C. C. J. Bacteriol. 1976, 126, 1166-1172). See FIG. 16 illustrating a propylene oxide conversion process to pyruvic acid. Additional epoxide substrates for selection may also be identified if *E. coli* mutants capable of growing on their corresponding vicinal diols.

[0366] Results. Epoxides are known to be toxic to microbes due to alkylation of proteins and nucleic acids. The tolerance of *E. coli* cells to epoxide compounds have been evaluated. In our preliminary experiments, we evaluated the effect of different concentrations of glycidol and propylene oxide on the growth of an *E. coli* host. The results showed that *E. coli* can tolerate up to 10 mM of glycidol and propylene oxide. This concentration may be high enough for selections as it was found that the cells were able to grow with 5 mM glycerol provided extracellularly in the media as the sole carbon source.

[0367] In addition, a propane diol-utilizing *E. coli* mutant and an *E. coli* mutant that can utilize glycerol more efficiently than the wild-type have been obtained from the *E. coli* Genetic Stock Center (CGSC) at Yale University. These hosts have been further engineered to be ready for use in selections.

Capillary Array Systems

[0368] The invention provides systems and methods for selecting a nucleic acid encoding an enantioselective enzyme and selecting enantioselective enzymes using capillary arrays, such as GIGAMATRIX™, Diversa Corporation, San Diego, Calif. The capillary arrays of the invention provide a system and method for "retaining" or "holding" nucleic acid and polypeptide samples to be analyzed. The nucleic acid and polypeptide samples (including reagents, substrates, etc.) can be directly or indirectly "held" onto a capillary wall. The cells in the cell-based methods can be directly or indirectly "held" within a capillary lumen. In one aspect, the capillary array apparatus includes a plurality of capillaries formed into an array of adjacent capillaries, wherein each capillary comprises at least one wall defining a lumen for "retaining" a sample. The apparatus can further include interstitial material disposed between adjacent capillaries in the array, and one or more reference indicia formed within of the interstitial material. The capillary can be adapted for being bound in an array of capillaries; e.g., it can include a first wall defining a lumen for retaining the sample and a second wall formed of a filtering material. The filtering material can, e.g., filter excitation energy provided to the lumen to excite the sample.

[0369] In one aspect, the capillary array includes a plurality of individual capillaries comprising at least one outer wall defining a lumen. The outer wall of the capillary can be one or more walls fused together. Similarly, the wall can define a lumen that is cylindrical, square, hexagonal or any other geometric shape so long as the walls form a lumen for retention of a liquid or sample. The capillaries of the capillary array can be held together in close proximity to form a planar structure. The capillaries can be bound together, by being fused (e.g., where the capillaries are made of glass), glued, bonded, or clamped side-by-side. The capillary array can be formed of any number of individual capillaries, for example, a range from 100 to 4,000,000 capillaries. A capillary array can form a microtiter plate having about 100,000 or more individual capillaries bound together.

[0370] The capillaries can be formed with an aspect ratio of 50:1. In one aspect, each capillary has a length of approximately 10 mm, and an internal diameter of the lumen of approximately 200 μm. However, other aspect ratios are possible, and range from 10:1 to well over 1000:1. Accordingly, individual capillaries have an inner diameter that

e.g., U.S. Pat. Nos. 5,874,259; 6,277,621; 6,183,957. BACs are based on the *E. coli* F factor plasmid system and simple to manipulate and purify in microgram quantities. Because BAC plasmids are kept at one to two copies per cell, the problems of rearrangement observed with YACs, which can also be employed in the present methods, are eliminated; see, e.g., Asakawa (1997) Gene 69-79; Cao (1999) Genome Res. 9:763-774. P1 artificial chromosomes (PACs), bacteriophage P1-derived vectors are, e.g., described in Woon (1998) Genomics 50:306-316; Boren (1996) Genome Res. 6:1123-1130; Ioannou (1994) Nature Genet. 6:84-89; Reid (1997) Genomics 43:366-375; Nothwang (1997) Genomics 41:370-378; Kern (1997) Biotechniques 23:120-124). P1 is a bacteriophage that infects *E. coli* that can contain 75 to 100 Kb DNA inserts (see, e.g., Mejia (1997) Genome Res 7:179-186; Ioannou (1994) Nat Genet 6:84-89). PACs are screened in much the same way as lambda libraries. See also Ashworth (1995) Analytical Biochem. 224:564-571; Gingrich (1996) Genomics 32:65-74. Other cloning vehicles can also be used, for example, recombinant viruses; cosmids, plasmids or cDNAs; see, e.g., U.S. Pat. Nos. 5,501,979; 5,288,641; 5,266,489. These vectors can include marker genes, such as, e.g., luciferase and green fluorescent protein genes (see, e.g., Baker (1997) Nucleic Acids Res 25:1950-1956). Sequences, inserts, clones, vectors and the like can be isolated from any natural sources, obtained from such sources as ATCC or GenBank libraries or commercial sources, or prepared by synthetic or recombinant methods. The nucleic acids and libraries of the invention can also be inserted and expressed in "expression cassettes," which comprise sequences capable of affecting expression of a structural gene (i.e., a protein coding sequence, such as an enzyme to be selected) in a host compatible with such sequences. Expression cassettes can include at least a promoter operably linked with the polypeptide coding sequence; and, optionally, with other sequences, e.g., transcription termination signals. Additional factors necessary or helpful in effecting expression may also be used, e.g., enhancers. "Operably linked" as used herein refers to linkage of a promoter upstream from a DNA sequence such that the promoter mediates transcription of the DNA sequence. Thus, expression cassettes also include plasmids, expression vectors, recombinant viruses, any form of recombinant "naked DNA" vector, and the like. A "vector" comprises a nucleic acid that can infect, transfect, transiently or permanently transduce a cell. It will be recognized that a vector can be a naked nucleic acid, or a nucleic acid complexed with protein or lipid. The vector optionally comprises viral or bacterial nucleic acids and/or proteins, and/or membranes (e.g., a cell membrane, a viral lipid envelope, etc.). Vectors include, but are not limited to replicons (e.g., RNA replicons, bacteriophages) to which fragments of DNA may be attached and become replicated. Vectors thus include, but are not limited to RNA, autonomous self-replicating circular or linear DNA or RNA (e.g., plasmids, viruses, and the like, see, e.g., U.S. Pat. No. 5,217,879), and includes both the expression and non-expression plasmids. Where a recombinant microorganism or cell culture is described as hosting an "expression vector" this includes both extra-chromosomal circular and linear DNA and DNA that has been incorporated into the host chromosome(s). Where a vector is being maintained by a host cell, the vector may either be stably replicated by the cells during mitosis as an autonomous structure, or is incorporated within the host's genome.

**[0376]** The phrases "nucleic acid" or "nucleic acid sequence" as used herein refer to an oligonucleotide, nucleotide, polynucleotide, or to a fragment of any of these, to DNA or RNA (e.g., mRNA, rRNA, tRNA) of genomic or synthetic origin which may be single-stranded or double-stranded and may represent a sense or antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material, natural or synthetic in origin, including, e.g., iRNA, ribonucleoproteins (e.g., iRNPs). The term encompasses nucleic acids, i.e., oligonucleotides, containing known analogues of natural nucleotides. The term also encompasses nucleic-acid-like structures with synthetic backbones, see e.g., Mata (1997) Toxicol. Appl. Pharmacol. 144:189-197; Strauss-Soukup (1997) Biochemistry 36:8692-8698; Samstag (1996) Antisense Nucleic Acid Drug Dev 6:153-156. The phrases "nucleic acid" or "nucleic acid sequence" as used herein refer to an oligonucleotide, nucleotide, polynucleotide, or to a fragment of any of these, to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent a sense or antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material, natural or synthetic in origin.

**[0377]** All headings and subheading used herein are provided for the convenience of the reader and should not be construed to limit the invention.

**[0378]** As used herein and in the appended claims, the singular forms "a," "and," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a clone" includes a plurality of clones and reference to "the nucleic acid sequence" generally includes reference to one or more nucleic acid sequences and equivalents thereof known to those skilled in the art, and so forth.

**[0379]** Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which the invention belongs. Although any methods, devices and materials similar or equivalent to those described herein can be used in the practice or testing of the invention, the preferred methods, devices and materials described.

**[0380]** All publications mentioned herein are incorporated herein by reference in full for the purpose of describing and disclosing the databases, proteins, and methodologies, which are described in the publications, which might be used in connection with the described invention. The publications discussed above and throughout the text are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention.

**[0381]** The invention will be further described with reference to the following examples; however, it is to be understood that the invention is not limited to such examples.

## EXAMPLES

### Example 1

### Selection of Enzymes which can Hydrolyze Individual Enantiomers of a β-hydroxynitrile

**[0382]** In the present example, the individual enantiomers of a β-hydroxynitrile are used as substrates for selection of

and 300 μl X-GAL (350 mg/ml)) are added and plated on 100 mm plates. The plates are incubated overnight at 37° C.

[0394] Amplification of Libraries ($5.0\times10^5$ recombinants from each library). About 3.0 ml host cells ($OD_{600}$=1.0) are added to two 50 ml conical tubes, inoculated with $2.5\times10^5$ pfu of phage per conical tube, and then incubated at 37°C. for 20 minutes. Top agar is added to each tube to a final volume of 45 ml. Each tube is plated across five 150 mm plates. The plates are incubated at 37° C. for 6-8 hours or until plaques are about pin-head in size. The plates are overlaid with 8-10 ml SM Buffer and placed at 4° C. overnight (with gentle rocking if possible).

[0395] Harvest Phage. The phage suspension is recovered by pouring the SM buffer off each plate into a 50 ml conical tube. About 3 ml of chloroform are added, shaken vigorously and incubated at room temperature for 15 minutes. The tubes are centrifuged at 2K rpm for 10 minutes to remove cell debris. The supernatant is poured into a sterile flask, 500 μl chloroform are added and stored at 4° C.

[0396] Titer Amplified Library. Serial dilutions of the harvested phage are made (for example, $10^{-5}$=1 μl amplified phage in 1 ml SM Buffer; $10^{-6}$=1 μl of the $10^{-3}$ dilution in 1 ml SM Buffer and the like), and 200 μl host (in 10 mM $MgSO_4$) are added to two tubes. One tube is inoculated with 10 μl of $10^{-6}$ dilution ($10^{-5}$). The other tube is inoculated with 1 μl of $10^{-6}$ dilution ($10^{-6}$), and incubated at 37° C. for 15 minutes.

[0397] About 3 ml of 48° C. top agar (50 ml stock containing 150 μl IPTG (0.5 M) and 37 μl X-GAL (350 mg/ml)) are added to each tube and plated on 100 mm plates. The plates are incubated overnight at 37° C.

[0398] The ZAP II library is excised to create the pBLUESCRIPT library according to manufacturer's protocols (Stratagene).

[0399] The DNA library can be transformed into host cells (e.g., *E. coli*) to generate an expression library of clones.

## Example 3

## Normalization

[0400] Prior to library generation, purified DNA can be normalized. DNA is first fractionated according to the following protocol A sample composed of genomic DNA is purified on a cesium-chloride gradient. The cesium chloride (Rf=1.3980) solution is filtered through a 0.2 μm filter and 15 ml is loaded into a 35 ml OptiSeal tube (Beckman) The DNA is added and thoroughly mixed. Ten micrograms of bis-benzimide (Sigma; Hoechst 33258) is added and mixed thoroughly. The tube is then filled with the filtered cesium chloride solution and spun in a Bti50 rotor in a Beckman L8-70 Ultracentrifuge at 33k rpm for 72 hours. Following centrifugation, a syringe pump and fractionator (Brandel Model 186) are used to drive the gradient through an ISCO UA-5UV absorbance detector set to 280 nm. Peaks representing the DNA from the organisms present in an environmental sample are obtained. Eubacterial sequences can be detected by PCR amplification of DNA encoding rRNA from a 10 fold dilution of the *E. coli* peak using the following primers to amplify:

```
Forward primer:
5'-AGAGTTTGATCCTGGCTCAG-3'        (SEQ ID NO:4)

Reverse primer:
5'-GGTTACCTTGTTACGACTT-3'         (SEQ ID NO:5)
```

[0401] Recovered DNA is sheared or enzymatically digested to 3-6 kb fragments. Lone-linker primers are ligated and the DNA is size-selected. Size-selected DNA is amplified by PCR, if necessary.

[0402] Normalization is then accomplished by resuspending the double-stranded DNA sample in hybridization buffer (0.12 M $NaH_2PO_4$, pH 6.8/0..82 M NaCl/1 mM EDTA/0.1% SDS). The sample is overlaid with mineral oil and denatured by boiling for 10 minutes. The sample is incubated at 68° C. for 12-36 hours. Double-stranded DNA is separated from single-stranded DNA according to standard protocols (Sambrook, 1989) on hydroxylapatite at 60° C. The single-stranded DNA fraction is desalted and amplified by PCR. The process is repeated for several more rounds (up to 5 or more).

## Example 4

## Construction of a Stable, Large Insert DNA Library of Picoplankton Genomic DNA

[0403] Cell collection and preparation of DNA. Agarose plugs containing concentrated picoplankton cells were prepared from samples collected on an oceanographic cruise from Newport, Oregon to Honolulu, Hi. Seawater (30 liters) was collected in Niskin bottles, screened through 10 μm Nitex, and concentrated by hollow fiber filtration (Amicon DC10) through 30,000 MW cutoff polyfulfone filters. The concentrated bacterioplankton cells were collected on a 0.22 μm, 47 mm Durapore filter, and resuspended in 1 ml of 2xSTE buffer (1 M NaCl, 0.1M EDTA, 10 mM Tris, pH 8.0) to a final density of approximately $1\times10^{10}$ cells per ml. The cell suspension was mixed with one volume of 1% molten Seaplaque LMP agarose (FMC) cooled to 40° C., and then immediately drawn into a 1 ml syringe. The syringe was sealed with parafilm and placed on ice for 10 min. The cell-containing agarose plug was extruded into 10 ml of Lysis Buffer (10 mM Tris pH 8.0, 50 mM NaCl, 0.1M EDTA, 1% Sarkosyl, 0.2% sodium deoxycholate, 1 mg/ml lysozyme) and incubated at 37° C. for one hour. The agarose plug was then transferred to 40 mls of ESP Buffer (1% Sarkosyl, 1 mg/ml proteinase K, in 0.5M EDTA), and incubated at 55° C. for 16 hours. The solution was decanted and replaced with fresh ESP Buffer, and incubated at 55° C. for an additional hour. The agarose plugs were then placed in 50 mM EDTA and stored at 4° C. shipboard for the duration of the oceanographic cruise.

[0404] One slice of an agarose plug (72 μl) prepared from a sample collected off the Oregon coast was dialyzed overnight at 4° C. against 1 mL of buffer A (100 mM NaCl, 10 mM Bis Tris Propane-HCl, 100 μg/ml acetylated BSA: pH 7.0 at 25° C.) in a 2 mL microcentrifuge tube. The solution was replaced with 250 μl of fresh buffer A containing 10 mM $MgCl_2$ and 1 mM DTT and incubated on a rocking platform for 1 hr at room temperature. The solution was then changed to 250 μl of the same buffer containing 4U of Sau3A1 (NEB), equilibrated to 37° C. in a water bath, and then

the product is shifted in the direction of product formation by addition of an excess of substrate or precursor.

8. The method of claim 1 or claim 2, wherein the equilibrium of the conversion of the product to the substrate or precursor is shifted in the direction of substrate or precursor formation by addition of an excess of product.

9. The method of claim 1 or claim 2, wherein the enzyme is a transaminase.

10. The method of claim 9, wherein the substrate or precursor is a specific enantiomer of an amino acid and the product is a specific enantiomer of an amino donor.

11. The method of claim 10, wherein the product further comprises an α-keto acid.

12. The method of claim 9, wherein the substrate or precursor comprises a specific enantiomer of an amino donor and the product is a specific enantiomer of an amino acid.

13. The method of claim 12, wherein the equilibrium of the conversion is shifted in the direction of amino acid product formation by addition of excess amino donor.

14. The method of claim 13, further comprising adding an α-keto acid amino acceptor to the media.

15. The method of claim 10, wherein the equilibrium of the conversion of the product to the substrate or precursor is shifted in the direction of product formation by enzymatic removal of an α-keto acid product.

16. The method of claim 10, wherein the equilibrium of the conversion of the specific enantiomer to the specific growth source is shifted in the direction of product formation by chemical removal of α-keto acid product.

17. The method of claim 1 or claim 2, wherein the enzyme is a nitrilase, wherein by supplying only nitrile groups of a desired chirality enantioselective nitrilase enzymes are identified.

18. The method of claim 17, wherein the substrate or precursor comprises a specific enantiomer of a nitrile-containing compound and the product comprises a specific enantiomer of a corresponding carboxylic acid and ammonia.

19. The method of claim 17, wherein the substrate or precursor comprises a specific enantiomer of a carboxylic acid and the product comprises a specific enantiomer of a nitrile-containing compound.

20. The method of claim 17, wherein medium comprises a nitrogen-free minimal media for cell growth, thereby only clones that can hydrolyze a nitrile group will produce the nitrogen source required to grow.

21. The method of claim 1 or claim 2, wherein the enzyme is an aldolase.

22. The method of claim 21, wherein the enzyme is an aldolase and deoxyribose-5-phosphate comprises the substrate or precursor and acetaldehyde comprises a product of the reaction.

23. The method of claim 1 or claim 2, wherein the enzyme is an epoxide hydrolase.

24. The method of claim 23, wherein the substrate or precursor comprises a chiral or racemic epoxide that is hydrolyzed to a diol that can be used as a growth source.

25. The method of claim 23, wherein the substrate or precursor comprises a chiral or racemic glycidol.

26. The method of claim 23, wherein the substrate or precursor comprises a chiral or racemic propylene oxide.

27. The method of claim 21, wherein the product further comprises a glyceraldehyde-3-phosphate.

28. The method of claim 1 or claim 2, wherein the enzyme is an aldolase and deoxyribose, 5-O-methyl-deoxyribose, or dideoxyribose comprises the substrate or precursor and acetaldehyde comprises a product of the reaction.

29. The method of claim 28, wherein the precursor or substrate comprises a deoxyribose and the product further comprises a glyceraldehyde.

30. The method of claim 28, wherein the precursor or substrate comprises a 5-O-methyl-deoxyribose and the product further comprises a 3-O-methyl-deoxyribose.

31. The method of claim 28, wherein the precursor or substrate comprises a dideoxyribose and the product further comprises a lactaldehyde.

32. The method of claim 1 or claim 2, wherein the nucleic acid further comprises an expression cassette, an expression vector, a phage or a plasmid.

33. The method of claim 32, wherein the vector comprises a PAC, a BAC, a MAC or a YAC.

34. The method of claim 2, wherein the nucleic acid library comprises phagemid library cells.

35. The method of claim 1 or claim 2, wherein the medium comprises a solid substrate or a liquid media.

36. The method of claim 1 or claim 2, wherein the cells that cannot make a factor, element or composition essential for growth are auxotrophs.

37. The method of claim 36, wherein the auxotrophs are amino acid auxotrophs.

38. The method of claim 37, wherein the auxotrophs can be made by a knockout strategy.

39. The method of claim 38, wherein the knockout strategy comprises a transposon mutagenesis.

40. The method of claim 1 or claim 2, wherein inserting the nucleic acid into a cell comprises infecting, transducing, transforming or transfecting the cells with the nucleic acid.

41. The method of claim 1 or claim 2, wherein the cells comprise bacterial cells.

42. The method of claim 41, wherein the bacterial cells comprise *E. coli*, Streptomyces, or *Bacillus subtilis*.

43. The method of claim 1 or claim 2, wherein the cells comprise fungal cells.

44. The method of claim 43, wherein the fungal cell comprise Aspergillus.

45. The method of claim 1 or claim 2, wherein the cells comprise insect cells.

46. The method of claim 45, wherein the insect cells comprise Drosophila S2 or Spodoptera Sf9.

47. The method of claim 1 or claim 2, wherein the cells comprise animal cells.

48. The method of claim 47, wherein the animal cells comprise a CHO cell, a COS cell or a Bowes melanoma cell.

49. The method of claim 1 or claim 2, wherein the cells comprise plant cells.

50. A method for identifying an enantioselective enzyme comprising the following steps:

(a) providing a polypeptide;

(b) providing a plurality of cells, wherein the cells cannot make a factor or element essential for growth and the essential factor or element must be of a specific chirality to induce growth of the cell;

(c) providing a substrate of a specific chirality, wherein the substrate is capable of being converted to an essential factor or element of the same chirality by an enzyme;

(f) selecting a sample producing the enzyme's chiral reaction product and identifying the added polypeptide of step (d), wherein selecting the sample comprising the enzyme's chiral reaction product selects an enzyme with a specific stereoselectivity profile.

64. The method of claim 63, wherein the polypeptide comprises a crude, partially purified, or purified enzyme.

65. The method of claim 62 or claim 63, further comprising immobilizing the polypeptide or the nucleic acid.

66. The method of claim 62 or claim 63, wherein the enzymatic reaction takes place on a substrate surface.

67. The method of claim 62 or claim 63, wherein the enzymatic reaction takes place in a capillary tube.

68. The method of claim 62 or claim 63, wherein the enzymatic reaction takes place in a double-orificed capillary array.

69. The method of claim 68, wherein the double-orificed capillary array is a GIGAMATRIX™ capillary array.

70. The method of claim 1, claim 2, claim 62 or claim 63, wherein combinations of enzymes are used.

71. The method of claim 1 or claim 2, wherein the method takes place in a double-orificed capillary array.

72. The method of claim 71, wherein the double-orificed capillary array is a GIGAMATRIX™ capillary array.

73. A method for selecting a nucleic acid encoding an enantioselective enzyme comprising the following steps:

(a) providing a nucleic acid encoding a polypeptide;

(b) providing a plurality of cells, wherein the cells cannot make a factor or element essential for growth and the essential factor or element must be of a specific chirality to induce growth of the cell;

(c) providing a substrate of a specific chirality, wherein the substrate is capable of being converted to an essential factor or element of the same chirality by an enzyme;

(d) inserting the nucleic acid into the cells and growing the cells under conditions wherein the nucleic acid is expressed and its encoded polypeptide is translated, and the cells are grown in a medium lacking the factor or element essential for growth, and adding the substrate of step (c); and,

(e) screening the cells for growth, wherein the nucleic acid in the growth stimulated clone is identified as encoding an enantioselective enzyme capable of converting the chiral substrate to a product comprising the chiral essential factor or element and this essential factor or element product of the enzyme's reaction is a single enantiomer, thereby selecting a nucleic acid encoding an enantioselective enzyme.

74. A method for identifying a nucleic acid encoding an enantioselective enzyme comprising the following steps:

(a) providing a nucleic acid library;

(b) providing a precursor of a specific chirality for a composition essential for growth, wherein the precursor is capable of being enzymatically converted to a product comprising the composition essential for growth, and to be growth-stimulating the composition essential for growth must have a chirality corresponding to the chirality of the precursor;

(c) providing a plurality of cells, wherein the cells cannot make the composition essential for growth;

(d) inserting in a cell a member of the gene library and culturing the cells in a medium lacking the composition essential for growth;

(e) adding the precursor of step (b) to the culture;

(f) selecting a growing cell and identifying the inserted library member of step (d), wherein the cell is capable of growth by enzymatic conversion of the precursor to a product comprising the composition essential for growth, and the enzyme is encoded by the library member, thereby identifying a nucleic acid encoding an enantioselective enzyme.

* * * * *